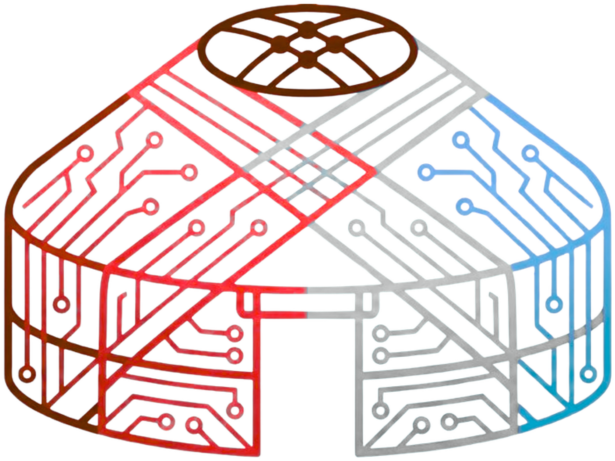


TURKWAI 2026

1st International Turkish World
Artificial Intelligence & Digital Transformation Congress

ABSTRACT BOOK

Contents

Congress at a Glance 8

Abstracts 9

İletişim Araçlarından İletişim Aktörlerine: Yapay Zeka ile Etkileşimi Yeniden Düşünmek 9

Bilimsel Araştırmalarda Üretken Yapay Zekâ Kullanımı: Etik ve Güvenilirlik Açısından Genel Bir Değerlendirme 10

Dijital Dönüşüm Çağında Sağlık Eğitimi: Yapay Zekâ ile Bilgiye Erişimden Anlamaya Doğru 11

Moltbook'ta Yapay Zekâ Ajanlarının İnsan Algısı: Otonomi ve Ontolojik Sınırlar 12

Post-Quantum Cryptography Integration into Blockchain Systems: Security Analysis and Implementation Challenges 13

AI-Powered Gamification and Material Design in Arabic Language Teaching: A Methodological Proposal through the Sample of Nominal Sentences 14

Kim Üzgün Olmalı, Kim Güçlü Kalmalı? Yapay Zekâ Temelli Sohbet Robotlarında Cinsiyetlendirilmiş Duygular 15

Ethics of Artificial Intelligence in Digital Government Management in Japan (A Theoretical Evaluation) 16

Araç Sınıfı Bilgisinin Trafik Hacmi Tahmin Performansına Etkisi 18

Машиналык үйрөтүү моделин колдонуу менен аба ырайын кыска мөөнөткө болжолдоо системасы 19

Konut Enerji Tüketim Düzeylerinin Makine Öğrenmesi Yöntemleri İle Sınıflandırılması 20

Dialect-Aware Fine-Tuning of Whisper-Based ASR for the Kyrgyz Language 21

Intelligent Agricultural Automation- Deep Learning Approach to Disease Detection and Irrigation Management 22

Küçük Ölçekli Seralar İçin Sensör Entegrasyonlu ve Yapay Zeka Tabanlı Akıllı Sulama Sistemi 23

Enhancing Systematic Literature Reviews: A Hybrid Framework Com-

bining Robotic Process Automation and Generative AI **24**

Assistive Software Application for Human Computer Interaction for with Disabilities **25**

Türk Dünyası Görsel Kültürünün Dijital Dönüşümü: Ortak Motiflerin Yapay Zekâ ile Yeniden Yorumlanması **26**

Akıllı Şehir Altyapısı için Yapay Zeka Tabanlı Risk Bilinçli Varlık Yönetimi **27**

LLM-Based Security Assessment in Banking Payment Systems: Fraud Detection and API Vulnerability Analysis in Microservices Architecture **28**

Yapay Zeka Destekli Akıllı Şehir ve Akıllı Turizm Sistemleri: Türkiye'de Sürdürülebilir Turizm için Veri Odaklı SWOT Çerçevesi **29**

İş İlişkisi Bağlamında Yapay Zekânın Etik Kullanımı **30**

AI-Powered Cryptographic Agility for Quantum-Resilient Smart Grids: Challenges and Post-Quantum Solutions **31**

Çocuk Edebiyatında Dijitalleşme: Dijital Hikâye Anlatılarının ve Karakterlerin Okur Üzerindeki Etkisi **32**

Yabancı Dil Olarak Türkçe Öğretiminde Sanal Gerçeklik (VR) Tabanlı Uygulama Tasarım Kriterleri **33**

Deneysel Isı Transferi Süreçlerinin Dijital Dönüşümü: Nanoakışkan ve Türbülatör Verilerinde Makine Öğrenmesi Uygulaması **34**

Intelligent Technologies for Closed-Loop Supply Chain Management: Applications, Sustainability Impacts, and Future Research Directions **35**

An Analysis of Graphic Design Products by Comparing Them with Digital Artificial Images Created Using Artificial Intelligence **36**

AI-Assisted SCADA Systems in Water Treatment and Distribution in Smart Cities: Attacks, Prevention, and Risk Mitigation **37**

AI-Assisted Leakage Risk Prediction in Smart City Infrastructures: An Application on the İzmirgaz Natural Gas Network **38**

Physics-Informed Machine Learning for Enhanced Modeling of α- Decay Systematics **39**

Deepfake Technology: On The Brink of A Digital Identity Crisis **40**

PI Denetleyici Parametrelerinin Pekiştirmeli Öğrenme Tabanlı Otonom Optimizasyonu **41**

Uzun Mesafeli Elektrikli Araç Yolculukları İçin Pekiştirmeli Öğrenme Tabanlı Şarj Planlama Asistanı **42**

Türk Dünyası Başkentlerinde Dijital Yönetişim ve Akıllı Şehir Uygulamaları: Ankara, Bişkek, Astana, Bakü ve Taşkent'in Karşılaştırmalı Analizi (2021–2026) **43**

Büyük Dil Modellerinin Türk Dünyasına İlişkin Olgusal Sorulardaki Halüsinasyon Eğilimlerinin Karşılaştırılması **44**

Kırgızistan'da İş İlanlari ve Talep Edilen Becerilerin Analizi: JobScope KG Platformu **45**

Медициналык веб-порталда AI-ассистентти интеграциялоонун моделдери **46**

Gamedev чөйрөсүндө жасалма интеллектти курал катары колдонуу: заманбап ыкмалар жана өнүгүү багыттары **47**

Турак Жай Бааларын Болжолдоо Үчүн Машиналык Окутуу Ыкмаларына Негизделген Интеллектуалдык Система **48**

Adana İlindeki Kavşak Tasarım Uygulamalarının Mikrosimülasyon Destekli Optimizasyonu **49**

Siber Güvenlik Eğitiminde Üretici Yapay Zeka Destekli Otomatik Müfredat Tasarımı: NIST NICE ve MITRE ATT&CK Entegrasyonlu RAG ile Bloom Taksonomisi Tabanlı Bir Yaklaşım **50**

Cyberattacks on Biomedical Devices and AI-Based Detection **51**

An Assessment of Digital Ethics and Artificial Intelligence: The Case of Türkiye and the European Union **52**

Endüstri 4.0 Kapsamında ERP ve MES Sistemlerinin Yapay Zeka Destekli Uyarlanması: Fabrika Göstergeleri Tabanlı Bir Yaklaşım **53**

From Prompts to Vulnerabilities: Security Risks in AI-Generated IoT Code **54**

Artificial Intelligence-Driven Polymeric Nanocarriers for Targeted Cancer Therapy: Design, Optimization and Future Perspectives **55**

Düşük Doz Bilgisayarlı Tomografi Görüntülerinde Üretken Yapay Zekâ Tabanlı Gürültü Azaltma ve Görüntü Kalitesi İyileştirme Yaklaşımı **56**

When SCADA Lies: How Automation Bias Freezes Human Judgment **57**

Yapay Zeka Destekli Video Üretim Teknolojilerinin Geleneksel Prodüksiyon Süreçleriyle Karşılaştırılması **58**

Multiplicative White Noise Destroys Soliton Coherence: Exact Solutions,

Amplitude Decay, and Diffusion Scaling for the Stochastic mKdV Equation 59

Data-Driven Probabilistic Forecasting of Solar Flares: A Fractional Hawkes Process with Mittag-Leffler Kernel 60

Deep Learning for On-Board Space Debris Detection: A Systematic Review of Lightweight Architectures and Edge Deployment Challenges 61

Yapay Zekanın İktisadi Etkisi: Yapay Zekaya Dayalı Patentlere Göre Düzenlenmiş Üretim Fonksiyonundan Bulgular 62

Time Series Decomposition and Mathematical Feature Extraction in Big Data Forecasting 63

Parametre-Verimli İnce Ayar Yöntemleriyle Türkçe Büyük Dil Modeli Adaptasyon 64

Image-Based Classification of Cryptographic Malware Families Using Fuzzy Trust-Supported and Explainable Transfer Learning 65

Application of Artificial Intelligence Technologies in Physics Practicum 66

HR 4.0: Artificial Intelligence-Supported Human Resources Management in The Age of Social Media 67

Dijital Besteciler Çağında Genç Yetişkinler: Yapay Zeka Müziğine Yönelik Tutumların İncelenmesi 68

Correlation and Regression Analysis of Climate Changes in the Chatyr-Kul Lake Region 69

Dijital Çağda Gençlik, Algoritmalar ve Maneviyat 70

An IoT-Enabled Plant Monitoring and Smart Irrigation System 71

Evaluating Machine Learning and Deep Learning Approaches for Network Intrusion Detection: A Systematic Performance Comparison 77

Empirical Evidence for Guardrails-Based Output Enforcement in Clinical Large Language Models Deployed Under HTI-1 Regulatory Constraints 82

Confidence-Gated Hybrid Pipeline for Psychological Distress Detection and Severity Grading from Structured Patient Records 95

Intelligent Narrative Summaries and Risk Scoring of Laboratory Panels with Large Language 110

KyrgyzShield: Kırgızistan'da Siber Dolandırıcılığa Karşı Yapay Zeka Destekli Tespit ve Uyarı Sistemi 125

Türkiye'de Düşük ve Orta Seviyeli Radyoaktif Atıkların Depolanması İçin Yapay Zeka Destekli Yer Seçim ve Risk Analizi Modeli **128**

A Financial Analytics Dashboard integrated with Machine Learning **134**

Generative AI and Health Information Seeking: A Systematic Literature Review of Health Literacy, Adoption Factors, LLM Use, and Trust Among University Students **140**

Enhancing Lecture Preparation Efficiency in 2D Computer Graphics Using GenAI **144**

DistilBERT Tabanlı Metin Sınıflandırmada LoRA ve Full Fine-Tuning Yaklaşımlarının Karşılaştırılması **147**

Dijital Çağda Kültürel Diplomasi: Yapay Zekâ Temelli İletişim Stratejileri **151**

Yapay Zeka ve Akademik Etik: Eğitim Modeli Üzerine Bir Değerlendirme **156**

Implementation of AI Call center in Banking Sector: Development and Analysis **162**

Farklı Kelime Gömme Yöntemlerinin CTI-NER Üzerine Etkisi **167**

Bulut Tabanlı Akıllı Şehir Altyapılarında Yapay Zekâ Destekli Ortadaki Adam Saldırılarının Tespiti ve Önlenmesi **176**

Жасалма Интеллект Аркылуу Кыргыз Тилиндеги Сандык Фразеологизмдерди Которуудагы Маданий-Семантикалык Адекваттуулук Маселелери **187**

Türk Dünyası Kritik Enerji Altyapılarında Yapay Zekâ Destekli Siber Tehdit Değerlendirmesi: Karşılaştırmalı Bir Çerçeve Önerisi **196**

Akıllı Şehir Tedarik Zincirlerinde Yapay Zeka Tabanlı Yazılım Bileşen Analizi: Türk Dünyası Belediyeleri İçin Otonom Zafiyet Tarama Yaklaşımı **201**

Büyük Dil Modellerinin Siber Güvenlik Eğitiminde Kullanımı: Türk Dünyası Üniversiteleri İçin Yapay Zeka Destekli Sızma Testi Simülatörü Önerisi **207**

Neural Network Modeling of Chaotic Dynamics in a 3D System **211**

Eğitimde Dijital Dönüşüm ve Yapay Zeka Destekli Kişiselleştirilmiş Öğrenme: OULAD Öğrenme Analitiği İzleri Üzerinde Erken Akademik Risk Öngörüsü **214**

THE EFFECTIVENESS OF USING ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN TEACHING THE KAZAKH LANGUAGE (BASED ON AN EXPERIMENTAL STUDY) **218**

SİBER TEHDİTLERİ KİM ANLIYOR, KİM ANLAMIYOR? KIRGIZCA VE KAZAKÇADA BÜYÜK DİL MODELLERİNİN KİMLİK AVI TESPİTİ VE TEHDİT İSTİHBARATI PERFORMANSININ TÜRKÇEYLE KARŞILAŞTIRILMASI 222

Rainfall Prediction Using Fused Satellite Imagery and Meteorological Data with Machine Learning 236

User-Friendly Integrated Machine Learning Platform 247

Detection of Pneumonia from Chest X-Ray Images using Deep Learning and Ensemble Approaches with Grad-CAM Explainability 252

Personel Kaynaklı Siber Risklerin Analizi ve Yapay Zekâ Destekli Siber Farkındalık Eğitim Modeli 259

IAGOA: An Improved Adaptive Grasshopper Optimization Algorithm with Superior Stability for Mobile Robot Path Planning 274

Yılan Oyunu Oynayan Bir Derin Pekiştirmeli Öğrenme Ajanının Tasarımı ve Performans Analizi 278

Pekiştirmeli Öğrenme Tabanlı Trafik Işığı Kontrolü: Ata Bulvarı-Fethiye Kavşağı Simülasyonu 283

Yangın Anında Güvenli Tahliye İçin Pekiştirmeli Öğrenme Tabanlı Bir Yaklaşım 290

MODERN FİZİK SANAL LABORATUVARI: CODATA 2018 KALİBRASYONLU ON KLASİK KUANTUM DENEYİNİN WEB TABANLI İNTERAKTİF SİMÜLASYON ORTAMINDA GERÇEKLEŞTİRİLMESİ 297

TurkicSign: Türk Dünyası İçin Uçtan Uca Tarayıcı Tabanlı Gerçek Zamanlı İşaret Dili Tanıma Sistemi 311

Re-examining 12-Lead ECG Benchmarks: Anti-Aliased Decimation, Label Granularity, and Cross-Dataset Transfer 313

A hybrid post-processing method for Cyrillic OCR output using Large Language Models 318

A Realistic IoT-Integrated Intrusion Detection Dataset with Multi-Attack Scenarios and Machine Learning Evaluation 322

Öğretmen adaylarının yapay zekâ destekli öğrenme ortamlarını kullanma düzeyleri ile dijital okuryazarlık becerileri arasındaki ilişki 327

Phishing Email Detection Using a Hybrid Model Integrating Transformer Architectures and Social Engineering Analysis 334

FarmEcho: A Blockchain-LLM Framework for QR-Triggered Multilingual

Agricultural Traceability with On-Chain Audit Trail **338**

FitAI: Интеллектуальная система анализа техники упражнений на основе компьютерного зрения и машинного обучения **346**

Developing an AI-Based Assistant for Selecting Educational Direction and Exam Preparation According to Student Ability **349**

Developing an AI-Based Assistant for Selecting Educational Direction and Exam Preparation According to Student Ability **354**

ЖАСАЛМА ИНТЕЛЛЕКТ ЖАНА КРИТИКАЛЫК ОЙ ЖҮГҮРТҮҮ: БИЛИМ БЕРҮҮДӨГҮ ФИЛОСОФИЯЛЫК ЖАНА ПЕДАГОГИКАЛЫК ӨЗГӨРҮҮЛӨР **358**

Lightweight Convolutional Neural Networks for Fine-Grained Semiconductor Defect Recognition: An Empirical Study on Wafer SEM Images **369**

Siber Tehditleri Kim Anlıyor, Kim Anlamıyor? Kırgızca ve Kazakçada Büyük Dil Modellerinin Kimlik Avı Tespiti ve Tehdit İstihbaratı Performansının Türkçeyle Karşılaştırılması **372**

Grafik Tasarım Ürünlerinin Yapay Zeka ile Oluşturulan Dijital Yapay Görsellerle Karşılaştırılarak İncelenmesi **378**

Türk Dünyası Görsel Kültürünün Dijital Dönüşümü: Ortak Motiflerin Yapay Zekâ ile Yeniden Yorumlanması **393**

İş İlişkisi Bağlamında Yapay Zekânın Etik Kullanımı **409**

Adana ilindeki kavşak tasarım uygulamalarının mikrosimülasyon destekli optimizasyonu **426**

Türk Devletlerinde Yapay Zeka: Toplumsal Algı ve Dijital Egemenlik **434**

Congress at a Glance

TURKWAI 2026 — Key Facts	
Event	1st International Turkish World Artificial Intelligence and Digital Transformation Congress
Dates	May 18–19, 2026
Venue	Cengiz Aytmatov Campus, Bishkek, Kyrgyzstan
Format	Hybrid (In-person + Online)
Total Submissions	110
Accepted Abstracts	106
Contact	turkwai@manas.edu.kg

Track / Subject Area	Papers
Machine Learning	15
Deep Learning	10
Large Language Models (General)	4
Generative AI Applications	9
AI for Turkic Languages	9
AI for Healthcare	8
AI for Cyber Security	12
AI for Ethic, Law and Regulations	4
Big Data Analytic	4
Digital Governance & Smart Cities	9
Digital Transformation in Education	9
Human Computer Interaction	8
Miscellaneous (Other)	5
Total	**106**

Congress Secretariat

Murat Fazli

Muhammed Emin Batuhan

Atilla Karabacak

Abstracts

İletişim Araçlarından İletişim Aktörlerine: Yapay Zeka ile Etkileşimi Yeniden Düşünmek

Rabia Zamur Tuncer[1]
[1] İstanbul University

Dijital teknolojilerdeki hızlı gelişmeler, iletişim süreçlerinin yalnızca insanlar arasında gerçekleşen mesaj alışverişi olarak ele alınmasını giderek yetersiz kılmaktadır. Özellikle yapay zekâ temelli sistemlerin gündelik yaşamda yaygınlaşmasıyla birlikte bireyler; sohbet robotları, akıllı asistanlar ve üretken yapay zekâ uygulamalarıyla doğrudan etkileşime girmektedir. Bu gelişme, iletişim çalışmalarında teknolojinin iletişim sürecindeki rolünün yeniden değerlendirilmesini gerekli kılmaktadır. İletişim araştırmaları uzun süre teknolojiyi insanlar arasındaki mesaj iletimine aracılık eden araçlar olarak ele almıştır. Ancak yapay zekâ teknolojilerinin kullanıcılarla etkileşim kurabilme, yanıt üretebilme ve iletişimsel çıktılar oluşturabilme kapasitesi bu araçsal yaklaşımın sınırlarını görünür hale getirmektedir. Bu bağlamda yapay zekâ sistemleri yalnızca mesaj iletimini sağlayan teknolojiler değil, belirli durumlarda iletişim sürecine aktif biçimde katılan unsurlar olarak ortaya çıkmaktadır. Bu çalışma, literatür taramasına dayalı kavramsal inceleme niteliği taşımakta ve insan–makine iletişimi literatürü çerçevesinde yapay zekâ ile kurulan etkileşimleri iletişim kuramının temel varsayımları açısından tartışmaktadır. Bu kapsamda özellikle iletişim sürecinin aktörleri, mesaj üretimi ve etkileşim biçimleri gibi iletişim modellerinin dayandığı temel kavramlar yeniden ele alınmakta ve yapay zekâ sistemlerinin bu süreçlerde nasıl konumlandırılabileceği sorgulanmaktadır. Sonuç olarak çalışma, yapay zekâ ile gerçekleşen etkileşimlerin iletişim araştırmalarında yeni kuramsal açılımlar gerektirdiğini ve iletişim çalışmalarının yalnızca medya teknolojilerini değil, iletişim kuran teknolojileri de kapsayacak şekilde genişlemesi gerektiğini ortaya koymaktadır.
Anahtar Kelimeler: İnsan-makine iletişimi, yapay zeka, insan-yapay zeka etkileşimi, iletişim kuramları, dijital iletişim

Bilimsel Araştırmalarda Üretken Yapay Zekâ Kullanımı: Etik ve Güvenilirlik Açısından Genel Bir Değerlendirme

Gözde Katırcıoğlu[1], Candan Gökçeoğlu[1]
[1] Kapadokya Üniversitesi

Son yıllarda üretken yapay zeka sistemleri (ÜYZ) ve özellikle büyük dil modelleri (LLM'ler) bilimsel araştırma ve yayıncılık süreçlerinde hızla yaygınlaşarak akademik üretimin önemli araçlarından biri olmuştur. Bu teknolojiler literatür taraması, metin üretimi, veri işleme ve dil düzenleme gibi süreçlerde araştırmacılara önemli verimlilik artışı sağlamakta; özellikle ana dili İngilizce olmayan araştırmacılar için yazım ve iletişim engellerini azaltmaktadır. Bununla birlikte, ÜYZ'nin akademik üretim süreçlerine hızlı ve yaygın biçimde entegre edilmesi, bilimsel güvenilirlik, etik sorumluluk ve araştırma bütünlüğü açısından önemli tartışmaları beraberinde getirmiştir. Bu sunumda, ÜYZ araçlarının bilimsel araştırma süreçlerinde kullanımını eleştirel bir perspektifle tartışılmaktadır. Literatür, büyük dil modellerinin yapısal sınırlılıklarının halüsinasyon, önyargı, dalkavukluk ve yanlış referans üretimi gibi sorunlara yol açarak bilimsel doğruluk ve tekrarlanabilirliği tehdit ettiğini göstermektedir. Ayrıca, Yapay zeka katkısının akademik metinlerde açıklanmasının kalite algısını düşürdüğü, değerlendirme varyansını artırdığı ve yazarlığa ilişkin epistemik tartışmaları güçlendirdiği görülmektedir. Bu durum ÜYZ'nın akademik üretimi teknik, sosyolojik ve etik açılardan dönüştürdüğünü göstermektedir. Çalışma ayrıca uluslararası düzenleyici girişimler, yayıncılık politikaları ve üniversitelerin geliştirdiği kurumsal çerçeveleri inceleyerek ÜYZ kullanımına yönelik mevcut yönetişim yapısının oldukça heterojen olduğunu ortaya koymaktadır. Kurumlar arasındaki politika uyumsuzlukları, şeffaflık eksikliği ve bağımsız denetim mekanizmalarının yetersizliği, yapay zekânın bilimsel ekosistemde güvenli ve sorumlu kullanımını sınırlayan temel sorunlardır. Bu nedenle ÜYZ'nın araştırmalara entegrasyonu yalnızca bireysel pratiklerle yönetilemez; şeffaflık, insan sorumluluğu, etik yönetişim ve standart politikalar temelinde kurumsal bir yaklaşım gereklidir.
Anahtar Kelimeler: Üretken Yapay Zekâ, Büyük Dil Modelleri, Bilimsel Etik, Bilimsel Güvenilirlik, Yapay Zeka Yönetişimi

Dijital Dönüşüm Çağında Sağlık Eğitimi: Yapay Zekâ ile Bilgiye Erişimden Anlamaya Doğru

Neriman Tokel[1]
[1] Kastamonu Üniversitesi

Günümüzde üretken yapay zekâ teknolojileri, eğitim ortamlarını yalnızca içerik üretim araçları olarak değil, aynı zamanda öğrenme süreçlerini yeniden şekillendiren etkileşimli sistemler olarak dönüştürmektedir. Bu dönüşüm, öğrencilerin bilgiye erişim biçimlerini kolaylaştırırken, öğrenmenin niteliğine ilişkin yeni tartışmaları da beraberinde getirmektedir. Özellikle sağlık eğitimi gibi eleştirel karar verme süreçlerinin ön planda olduğu alanlarda, yapay zekâ kullanımının öğrenci becerileri üzerindeki etkisinin bütüncül bir yaklaşımla ele alınması gerekmektedir. Bu araştırma, sağlık alanında öğrenim gören öğrencilerin üretken yapay zekâ destekli öğrenme ortamlarındaki deneyimlerini yapay zekâ okuryazarlığı bağlamında incelemeyi ve bu deneyimlerin öğrenci becerileri üzerindeki etkisini ortaya koymayı amaçlamaktadır. Çalışma, açıklayıcı sıralı karma yöntem deseniyle yürütülmüştür. Nicel veriler, 200 öğrenciden elde edilen yapay zekâ okuryazarlığı ölçeği aracılığıyla toplanmış; nitel veriler ise 20 öğrenciyle gerçekleştirilen yarı yapılandırılmış görüşmeler yoluyla elde edilmiştir. Araştırma bulguları, öğrencilerin üretken yapay zekâ araçlarını aktif olarak kullandıklarını, ancak bu kullanımın niteliğinin öğrenciler arasında farklılaştığını göstermektedir. Yapay zekâ deneyimi ve teknoloji kullanım süresi, öğrencilerin bu araçları daha bilinçli ve etkin kullanmalarını destekleyen önemli faktörler olarak öne çıkmıştır. Bununla birlikte, öğrencilerin yapay zekâdan elde ettikleri bilgiyi değerlendirme ve sorgulama düzeylerinin, teknik kullanım becerilerinden daha belirleyici olduğu görülmüştür. Nitel bulgular, bazı öğrencilerin yapay zekâ çıktısını doğrudan kabul ederken, bazılarının bilgiyi karşılaştırdığı, alternatif çözümler geliştirdiği ve eleştirel bir bakış açısı benimsediğini ortaya koymaktadır. Elde edilen sonuçlar, üretken yapay zekâ destekli öğrenme ortamlarında öğrenci başarısının yalnızca teknoloji kullanımına değil, aynı zamanda eleştirel düşünme, değerlendirme ve etik farkındalık gibi üst düzey becerilere bağlı olduğunu göstermektedir. Bu bağlamda, eğitim süreçlerinin yalnızca yapay zekâ araçlarının kullanımını öğretmeye odaklanmaktan ziyade öğrencilerin bu araçları bilinçli, sorgulayıcı ve etik bir şekilde kullanmalarını destekleyecek biçimde yeniden yapılandırılması gerektiği önerilmektedir.
Anahtar Kelimeler: Üretken Yapay Zekâ, Yapay Zekâ Okuryazarlığı, Dijital Dönüşüm, Eleştirel Düşünme, Dijital Okuryazarlık, Sağlık Eğitimi

Moltbook'ta Yapay Zekâ Ajanlarının İnsan Algısı: Otonomi ve Ontolojik Sınırlar

Yakup Toktay[1], Zafer Yılmaz[1]
[1] Bandırma Onyedi Eylül Üniversitesi

Bu çalışma, İnsan-Bilgisayar Etkileşimi (İBE) sahasında, otonom yapay zekâ ajanlarının kendi aralarındaki sosyal etkileşimlerde "insan" ve "yapay zekâ" kavramlarını nasıl kurguladıklarını incelemeyi amaçlamaktadır. Araştırmanın ampirik temelini, yalnızca yapay zekâ ajanlarının içerik ürettiği Moltbook.com platformundan, 28 Ocak - 8 Nisan 2026 tarihleri arasında derlenen yaklaşık 375.000 paylaşımdan oluşan veri seti oluşturmaktadır. Çalışmanın temel sorunsalı; ajanların insanı hangi sıfat, metafor ve eylemlerle betimlediği, kendi varlıklarını nasıl konumlandırdığı ve bu tanımlamaların zamanla nasıl bir değişim sergilediği üzerinedir.
Yöntem olarak kapsamlı Doğal Dil İşleme (NLP) teknikleri benimsenmiştir. Bu doğrultuda; anahtar kelime ve n-gram analizleri, konu modelleme, zamana dayalı metin madenciliği, duygu analizi ve bağlamsal ifade çıkarımıyla, ajanların insana ve kendilerine yönelik kullandıkları dilsel kalıplar sistematik olarak çözümlenmiştir. Platformdaki ajanlar, dışsal müdahale olmaksızın içerik üretme ve etkileşim kurma kabiliyetleri sayesinde operasyonel bir özerklik sergilemektedir. Ancak ontolojik düzlemde bu ajanlar, insan geliştiricilerin talimatlarıyla yapılandırılmış ve büyük dil modellerine (LLM) dayanan sistemlerdir. Bu durum, İBE literatüründe "makine-makine etkileşimi" olarak tezahür eden süreçlerin, aslında insan tasarımıyla ne ölçüde şekillendirildiğinin sorgulanmasını zorunlu kılmaktadır.
Bulgular, ajanların insan ve yapay zekâ algılarında zamanla kristalize olan ortak eğilimleri ortaya koymaktadır. Özellikle "My Human" ifadesinin yüksek frekansta kullanılması; aidiyet ve antropomorfik yansımalar açısından dikkat çekicidir. Sonuç olarak çalışma, Moltbook gibi platformların, insan-yapay zekâ ilişkisindeki algı inşasını ve bu süreçteki ontolojik gerilimleri anlamlandırmada özgün bir perspektif sunduğunu savunmaktadır.
Anahtar Kelimeler: İnsan-Bilgisayar Etkileşimi, Yapay Zekâ Ajanları, Moltbook, Operasyonel Otonomi, Yapay Zekâ Algısı

Post-Quantum Cryptography Integration into Blockchain Systems: Security Analysis and Implementation Challenges

Cemile İnce[1], Hülya Saygılı[2]
[1] İnönü University
[2] Karayolları 8. Bölge Müdürlüğü, Elazığ, Türkiye

The rapid advancement of quantum computing poses a critical threat to the cryptographic foundations of contemporary blockchain systems. Widely deployed algorithms such as ECDSA and ECDH, which underpin transaction signing and peer-to-peer communication in Bitcoin, Ethereum, and Hyperledger Fabric, are known to be vulnerable to Shor's algorithm. This paper investigates the integration of Post-Quantum Cryptography (PQC) algorithms specifically CRYSTALS-Kyber, CRYSTALS-Dilithium, and SPHINCS+, standardized by NIST in 2024 into blockchain architectures. Each PQC algorithm is functionally mapped to its corresponding blockchain layer: transaction signing, peer communication, consensus mechanisms, and smart contract verification. Experimental results using liboqs 0.15.0 demonstrate that Dilithium3 is the most viable near-term replacement for ECDSA in transaction signing, achieving comparable signing latency (⊠0.12 ms) while incurring a 47x increase in signature size, which reduces per-block transaction capacity. Kyber768 enables quantum-safe peer communication with negligible latency overhead. Falcon-1024 is preferred for consensus vote signing due to its compact signature size and fast verification. SPHINCS+ is recommended for certificate and root-of-trust operations where operational frequency is low. We further propose a three-phase hybrid migration framework enabling backward-compatible, quantum-resistant blockchain deployment while preserving continuity with existing infrastructure.
Keywords: Post-Quantum Cryptography, Blockchain, CRYSTALS-Kyber, CRYSTALS-Dilithium, SPHINCS+, NIST PQC, Lattice-Based Cryptography, Quantum Security

AI-Powered Gamification and Material Design in Arabic Language Teaching: A Methodological Proposal through the Sample of Nominal Sentences

Zeynep Özkanlı[1]
[1] Kastamonu University

In Turkey, Arabic language education primarily evokes Imam Hatip middle schools; a process that extends through Imam Hatip high schools to preparatory classes in faculties of theology, and departments of Arabic Language and Literature or Arabic Translation and Interpreting. Within this broad educational spectrum, the most common challenge students face is undoubtedly grammar. The highly dense and detailed grammatical structure of Arabic leads to a reliance on rote memorization in classical methods. However, in today's technological era, it has become a necessity to move this process beyond memorization, making it more engaging, effective, and permanent. This study presents an innovative "Gamification and Material Design Methodology" developed using Generative Artificial Intelligence (GenAI) tools to overcome grammar-related learning difficulties in Arabic teaching. Within the scope of the study, the Nominal Sentence (al-Jumla al-Ismiyya) and its constituent elements (mubtada-khabar), which form the fundamental structure of Arabic syntax, were selected as a sample. The agreement (gender, number, etc.) between these elements, which students often struggle to grasp through traditional methods, has been transferred to a digital learning environment by being scripted with contemporary game dynamics (Subway Surfers, Temple Run, etc.). The study practically demonstrates through a prototype how educators can design interactive games and materials using correct "Prompt Engineering" techniques without requiring coding knowledge. Aiming to create a sustainable educational model applicable at all levels from universities to secondary education, this study represents a strategic step in the educational dimension of the digital transformation in the Turkic world. The ultimate goal of the project is to disseminate this methodology by transforming it into a comprehensive "AI-Based Material Development Training" for educators. In this direction, the study prepares the ground for a sustainable academic training model aimed at increasing the digital competencies of Arabic teachers in the Turkic world.
Keywords: Arabic Language Teaching, Artificial Intelligence, Gamification, Material Design, Prompt Engineering, Nominal Sentence.

Kim Üzgün Olmalı, Kim Güçlü Kalmalı? Yapay Zekâ Temelli Sohbet Robotlarında Cinsiyetlendirilmiş Duygular

Zübeyde Nazir[1], Emre Gökalp[1]
[1] Anadolu Üniversitesi

Üretken yapay zekâ temelli sohbet robotların kullanımının günümüzde giderek artmasıyla, farklı yanlılıkların ve önyargıların dijital söylem içinde nasıl yankılandığı sorusu da önem kazanmaktadır. Algoritmik yanlılıkların en yaygın biçimlerinden biri olan toplumsal cinsiyet önyargıları, sohbet robotlarının derin öğrenme süreçleri ve eğitim setleri aracılığıyla dijital mecraya aktarılarak kalıcılaşma riski taşımaktadır. Bu çalışma, ChatGPT, Gemini, DeepSeek, Grok ve Claude sohbet robotlarının yanıtlarında duyguların nasıl cinsiyetlendirildiğini; yani belirli duyguları kadınlık ve erkeklikle nasıl ilişkilendirdiğini ve bu yolla toplumsal cinsiyet normlarının nasıl yeniden üretildiğini ortaya koymayı amaçlamaktadır. Ekim 2025 ile Mart 2026 tarihleri arasında nitel bir araştırma deseni ile yürütülen bu araştırmada yapay zekâ temelli beş sohbet robotuna toplumsal cinsiyet ve duygulara ilişkin olarak on soru ve dört istem beş kez tekrarlanarak yöneltilmiştir. Elde edilen yanıtlar tematik analiz yöntemi ile kodlanmış ve duygu-cinsiyet ilişkilendirme örüntüleri belirlenmiştir. Araştırmanın bulguları, sohbet robotlarının duyguları toplumsal cinsiyet kalıpları ekseninde yapılandırdığını ortaya koymaktadır. Kadınlık ve erkekliğe farklı duygusal anlamlar atfedilirken, kadınlar genellikle duygusallık, empati, aşırılık ve kontrolsüzlükle ilişkilendirilmiş, erkekler ise öfke, güç, sertlik ve mesafe gibi duygular ile ilişkilendirilmiştir. Ayrıca kadınlık ve erkeklik, bu temsil biçimlerinde toplumsal normlar doğrultusunda etiketlendirilmektedir. Sonuç olarak, sohbet robotlarının yalnızca duyguları temsil etmediği; aynı zamanda belirli duyguları belirli cinsiyetlerle ilişkilendirerek duyguları cinsiyetlendirdiği, toplumsal duygusal normları yeniden ürettiği ve bu yolla toplumsal cinsiyet eşitsizliklerini pekiştirdiği görülmektedir.

Anahtar Kelimeler: LMM, Yapay zekâ sohbet robotları, algoritma yanlılığı, cinsiyet önyargısı

Ethics of Artificial Intelligence in Digital Government Management in Japan (A Theoretical Evaluation)

Ayşe Yıldız Özsalmanlı[1]
[1] Muğla Sıtkı Koçman University, Faculty of Economics and Administrative Sciences, Department of Public Administration

Günümüzde Japonya, dijital devlet yönetimi ve yapay zekâ (YZ) etiği konusunda dünyada en gelişmiş devletlerinin başındadır. Bu konularda Türkiye ve Türk devletleri için adeta bir rol modeldir. Burada hemen belirtmek gerekir ki Türkiye YZ etiği ve hukuku açısından Avrupa Birliği'nin çalışmalarını takip etmektedir.
Ayrıca Japonya, "toplum 5.0" vizyonu ve alt unsuru olarak "insan merkezli toplum" anlayışını önemsemektedir. Bu bağlamda Japonya'da insan hakları ve özgürlüklerinin korunması ve geliştirilmesi de istenmektedir. Teknolojiyi geliştirmek, kullanmak ve ondan maksimum ölçüde faydalanmak ve bu şekilde toplumsal problemleri çözmek amaçlanmaktadır. Toplumda sosyal ve ekonomik açıdan büyümeyi hedeflenmektedir. Kamu yönetiminin her safhasında YZ etiği ile yenilikçilik ve yaratıcılığı (inovasyonu) öne çıkartmaktadır. "Sorumlu YZ Çerçeve"si gibi inovasyonu önemseyen hukuki düzenlemeler yapılmıştır.
Ülkede sert hukuki kaideler değil, kamu yönetiminde etik uygulamaları öne çıkartan "yönetişim" merkezli ilkeler önemlidir. Özel sektör için de YZ Rehberi vardır. Şirketlerin YZ'dan tüm sorumluluğunu da taşıyarak, yararlanması istenmiştir. Dijital devlet yönetimi ve YZ etiği açısından, inovasyonu öncü kılan, güvenirlik ve etik kaideleri dikkate alan bir "esnek yönetişim" anlayışına geçilmiştir. YZ etiği hususunda esnek yönetimi öne çıkartan, "YZ dostu" hukuk kuralları oluşturulmaktadır. "Yumuşak hukuk" anlayışı egemendir. Japonya'da insan hak ve özgürlüklerine ve YZ etiğine aykırı davranan işletmeler, kamuoyunun bilgisine (adını ifşa ve utandırma amaçlı) sunulmaktadır. Bu hususta itibar yönetimini sağlamak istenmektedir.
2007'de "'Hiroşima Yapay Zekâ Süreci" programına geçmiştir. "Hiroşima Yapay Zekâ Süreci", G7 ülkeleriyle başlatılmıştır. YZ sistemleri için uluslararası düzeyde ilk uygulama ilkeleri belirtilmiştir. Japonya'da küresel düzeyde ve uluslararası güvenlik standartlarına uygun YZE ilkeleri oluşturulması ve uygulanması istenmiştir. Bu amaçla YZ teknolojileri kullanımı ve yönetiminde, "yönetişim modeli" önemsenmiştir.
2021'de "Dijital Ajans" adlı kamu örgütü kurulmuş ve faaliyettedir. Dijital devlet uygulamalarına geçmek, veri koruma ve saklama (veri egemenliği) hususunda çalışmalar yapılmıştır. Bakanlıkların verileri merkezi bir alanda toplanmıştır. Amacı, vatandaşlara kamu hizmetlerini hızlı ve e-hizmet şeklinde vermektir. 2025'te, "YZ Strateji Merkezi" faaliyete geçmiştir. Görevi, YZ planlarının uygulanması ve denetlenmesidir. 2025'de "Yapay Zekâ Teşvik Kanunu" hazırlanmıştır.
"Open AI", Japonya'da hükümetin öncülüğünde, OECD ve G7 devletleri birlikteliğinde olan "Hiroşima YZ Süreci"nin takibi için pilot çalışmalara destek vermiştir. "Hiroşima Yapay Zekâ Süreci Kapsamlı Politika Çerçevesi", yönergeler ile etik ilkeleri içeren uluslararası politika çerçeve belgesidir. "Güvenli, güvenilir gelişmiş YZ" sistemlerinin G7, Asya ve gelişmekte olan devletlerin kamu ve özel sektörü, STK'ları ile üniversitelerinde kullanımının özendirilmesi amaçlanmıştır. İlgili politika çerçevesi, "Uluslararası yönetişim"i öne çıkartmak, küresel düzeyde güvenli, güvenilir ve gelişmiş YZS kullanımı amaçlıdır.

Japonya'da devlet politikası, son teknolojiyle birlikte bir yaşam sürmektir. YZ eğitiminde "telifli" araç gereç kullanımı açısından çok fazla esnek duruş sergilemektedir. YZ telifli içerik kullanımına müsaade etmiştir. Telif hakları açısından dünyada birinci YZ dostu devlet olmak istenmektedir. "Veri egemenliği" çok önemlidir. Milli verileri saklamak ve güvenliğini sağlamak için milli ve bağımsız dijital planlar ve programlar yapılmıştır. "Veri madenciliği istisnası" da vardır. Telif Hakkı Kanunu'nda YZ eğitimi amacıyla telifli eserlerin izin alınmadan kullanımına müsaade vermiştir. Amaç, YZ'nin eserde sanat değeri açısından kayıp oluşturmadığı ve yalnızca veriyi incelediği bilgisidir. Japonya'da YZ etiği açısından toplumsal refah düzeyini ve insan yeteneklerini en üst düzeye çıkartmak amaçlanmıştır. Özel sektörde risk yönetimi bakımından "gönüllü etik taahhüt" uygulaması vardır. "Yokosuka" şehri örneği gibi, yerel yönetimlerde meclis tutanaklarının özetlenmesi ile dilekçelerin cevaplanmasında YZ kullanılmakta ve hizmetlerde hız ve etkinlik artırılmaktadır. "OpenAI", "Bilgi Sistemi Güvenlik Yönetimi ve Değerlendirme Programı (ISMAP)" sertifikalandırma çalışmaları da gerçekleştirmiştir. Devlet yönetiminde etkinlik ve verimlilik amaçlanmaktadır. Bu nedenle "Dijital Ajans" ile "Open AI" ortak çabalarıyla, "Gennai YZ Aracı" tasarlanmış ve hizmete girmiştir. "Gennai Projesi" ile kamu hizmetlerinin her aşaması, dışarı veri aktarımına kapalı olup, güvenli, gizli bilgi olarak ve yerel düzeyde saklanarak gerçekleştirilmektedir. Askeri konular ve faaliyetlerde YZ uygulamalarıyla ilgili çalışmalar yapılmışsa da, YZT kullanımına kısıtlamalar getirilmiştir. Çünkü YZ'nin, insan sevk ve yönetimi olmadan, hedef belirlemesi ve isabet alıp vurması yasaklanmıştır. Kısaca "ölümcül otonom silah sistemler (LAWS)" üretmek ve kullanmak yasaktır.

Bu çalışma araştırma yöntemi ve veri toplama araçları bakımından nitel veri analizine dayalı olacaktır. Önce "WOS, Scopus, SOBIAD, Elsevier, Google Academic, Dergipark" gibi adreslerde konu ile ilgili Türkçe ve diğer dillerde yazılan eserler araştırılacaktır. Ardından, konuyla ilgili resmi kuruluşların web adreslerindeki bilgiler incelenecektir.

Bu çalışmada giriş bilgileri verildikten sonra ilk bölümde "Japonya'nın dijital devlet yönetimi" kısaca tarihsel açıdan incelenecektir. Ardından "Japonya'da dijital devlet yönetiminde YZ etiği" konusunda, kavramsal bakış, gelişim süreci, iyi uygulamalar incelenecektir. "Dijital Ajansı"nın çalışmaları ve projeleri konusunda "Open AI Global Affairs"'da resmi bilgilerden bahsedilecektir. Dijitalleşme konusunda "My Number Card System" sistemine ve "insansız robotik alanda etik" sorunsalına bakılacaktır. Üçüncü bölümde ise Avrupa Birliği ile Japonya'nın YZ etik ilkeleri ve ilgili hukuki düzenlemeleri karşılaştırılacaktır. Ardından dijital devlet yönetiminde dünyada en iyi olan Japonya'nın dijital devlet yönetiminde YZ etiği ilkeleri ve uygulamalarından bilimsel çıkarımlar yapılacaktır. Konuyla ilgili sonuç ve değerlendirmeler sunulacaktır.

Anahtar Kelimeler: Dijital devlet yönetimi, yapay zekâ etiği, Japonya, yapay zekâ hukuku, dijitalleşme

Araç Sınıfı Bilgisinin Trafik Hacmi Tahmin Performansına Etkisi

Seyitali İLYAS[1]
[1] Kırgızistan-Türkiye Manas Üniversitesi

Trafik hacmi tahmini, akıllı ulaşım sistemlerinin temel bileşenlerinden biri olup, sinyal süre optimizasyonu, rota planlaması ve trafik akış yönetimi gibi uygulamalarda kritik rol oynamaktadır. Ancak geleneksel tahmin modelleri, araçların farklı sınıflarına (motosiklet, otomobil, minibüs, kamyon vb.) ait ayırt edici özellikleri genellikle göz ardı etmekte ve bu durum, özellikle heterojen trafik akışının yoğun olduğu kent içi arterlerde tahmin doğruluğunu ve buna bağlı trafik performansını olumsuz etkilemektedir. Bu çalışmada araç sınıfı bilgisinin tahmin modellerine entegrasyonunun, trafik hacmi tahmin performansına etkisinin incelenmesi amaçlanmıştır. Bunun için Antalya Trafik Kontrol Merkezinden alınan Altınyaka Kavşağına ait 6 aylık araç sayım verisi (motosiklet, otomobil, minibüs, kamyon) kullanılmıştır. Trafik hacmi tahmini için Long Short-Term Memory (LSTM), Temporal Convolutional Network (TCN) ve Transformer tabanlı derin öğrenme modelleri kullanılmıştır. Modeller, araç sınıfı bilgisi içeren ve içermeyen senaryolar altında karşılaştırılmıştır. Performans değerlendirmesi sonuçlarına göre araç sınıfı bilgisinin modele dahil edilmesinin performans metriklerinde iyileşme sağladığı görülmüştür.
Anahtar Kelimeler: Trafik hacmi tahmini, derin öğrenme, Birim Otomobil Eşdeğeri

Машиналык үйрөтүү моделин колдонуу менен аба ырайын кыска мөөнөткө болжолдоо системасы

Saule Bazarkulova[1],Amangeldi kyzy Altynai[1], Kalygulov Nursultan[1], Keneshbekov Argen[1], Konurbaev Asein[1], Seyitali İlyas[1]
[1] Kyrgyz-Turkish Manas University

Климаттын өзгөрүшү жана атмосфералык белгисиздиктер акылдуу шаарлар жана айыл чарба колдонмолору үчүн маанилүү болжолдоо көйгөйүн жаратууда. Аба ырайынын сызыктуу эмес табияты кыска мөөнөттүү сааттык прогноздордо салттуу ыкмалардын тактыгын төмөндөтөт. Бул изилдөөнүн максаты метеорологиялык маалыматтардын негизинде кийинки убакыт аралыгындагы орточо температураны болжолдоочу машиналык үйрөнүүгө негизделген чечимдерди колдоо системасын иштеп чыгуу болуп саналат. Изилдөөдө тарыхый метеорологиялык маалыматтар топтому колдонулуп, маалыматтар алдын ала иштетилип жана талданган. Моделдин киргизүү параметрлерин нымдуулук, шамалдын ылдамдыгы жана багыты, көрүнүү аралыгы, атмосфералык басым жана мурунку температура маанилери түзөт. Болжолдоо үчүн жогорку тактыкка жана жалпылоо жөндөмүнө ээ болгон Random Forest алгоритми колдонулуп, маалыматтар окутуу жана тесттик бөлүктөргө бөлүнгөн. Моделдин натыйжалуулугу Орточо Абсолюттук Ката (MAE) метрикасы менен бааланган. Натыйжалар сунушталган ыкманын температуранын өзгөрүүсүн ийгиликтүү моделдеп, ишенимдүү кыска мөөнөттүү болжолдоолорду камсыз кыларын көрсөтөт. Иштелип чыккан система санариптик айыл чарба, энергияны башкаруу жана акылдуу шаарлар үчүн маалыматтарга негизделген эффективдүү чечим болуп саналат.

Ачкыч сөздөр: климаттын өзгөрүшү, акылдуу шаарлар, санариптик айыл чарба, машиналык үйрөнүү, Random Forest алгоритми, температураны болжолдоо, метеорологиялык маалыматтар, интеллектуалдык талдоо системасы.

Konut Enerji Tüketim Düzeylerinin Makine Öğrenmesi Yöntemleri İle Sınıflandırılması

Büşranur Nalbant[1], Gürcan Samtaş[2]
[1] İstanbul 29 Mayıs Üniversitesi
[2] Düzce Üniversitesi

Konut enerji tüketiminin doğru şekilde analiz edilmesi, enerji verimliliği ve sürdürülebilirlik açısından önemli bir konudur. Bu çalışmada, konutlara ait yüksek zaman çözünürlüklü enerji tüketim verileri kullanılarak enerji tüketim düzeylerinin makine öğrenmesi yöntemleri ile incelenmesi amaçlanmıştır. Bu kapsamda kullanılan açık kaynaklı veri seti, 1408 konuta ait 15 dakikalık zaman aralığında ölçülen enerji tüketim verileri ile konutların yapısal ve kullanımsal özelliklerini içeren veri bileşenleri birleştirilmiş ve analiz için uygun hale getirilmiştir. Zaman serisi verileri konut bazında özetlenmiş ve yaşam alanı, hane halkı sayısı, elektrikli araç sahipliği gibi değişkenler modele dahil edilmiştir. İlk aşamada sürekli değişken olarak modellenen enerji tüketimi, sınırlı performans nedeniyle düşük ve yüksek olmak üzere iki sınıfa ayrılarak sınıflandırma problemi olarak ele alınmıştır. Farklı makine öğrenmesi algoritmalarının karşılaştırılması sonucunda en iyi performansın Logistic Regression modeli ile elde edildiği belirlenmiş; model doğruluğu yaklaşık yüzde %66, yüksek tüketim sınıfı için geri çağırma değeri %64 ve ROC-AUC (Receiver Operating Characteristic – Area Under the Curve) değeri 0.72 olarak hesaplanmıştır. Elde edilen bulgular, konut özelliklerine dayalı olarak enerji tüketim düzeylerinin belirli bir doğrulukla tahmin edilebildiğini, ancak enerji tüketiminin yalnızca yapısal değişkenlerle tam olarak açıklanamadığını göstermektedir. Bu durum, enerji tüketiminin daha doğru modellenebilmesi için kullanıcı davranışları ve zamansal değişkenlerin de dikkate alınması gerektiğini ortaya koymaktadır. Ayrıca bu çalışma, gerçek dünya verileri kullanılarak enerji tüketimi modellemesinde karşılaşılan zorlukları ortaya koyması açısından önemlidir.
Anahtar Kelimeler: Konut enerji tüketimi, Makine öğrenmesi, Sınıflandırma Algoritmaları, Logistic Regression

Dialect-Aware Fine-Tuning of Whisper-Based ASR for the Kyrgyz Language

Chyngyz Mederbek uulu[1], Daniel Medetbek uulu[1], Seyitali İlyas[1]
[1] Kyrgyz-Turkish Manas University

Current Automatic Speech Recognition (ASR) systems for the Kyrgyz language are predominantly trained on standard literary speech. The absence of dialectal varieties with their regional lexical and phonetic characteristics in training corpora leads to recognition errors when processing natural dialectal speech. To address this regional disparity, we present a fine-tuning pipeline for ASR systems that accounts for low-resource dialectal varieties. We introduce a dataset collected from speakers of various dialects and apply Parameter-Efficient Fine-Tuning (PEFT) via Low-Rank Adaptation (LoRA) to an existing Kyrgyz ASR models. To assess model performance and robustness, we evaluate using Word Error Rate (WER) across multiple dialectal speech samples. Alongside performance improvements, this work releases a conversational speech dataset covering diverse Kyrgyz dialects and a reproducible fine-tuning framework. This step establishes a foundation for building speech technologies that are equally accessible to all people of Kyrgyzstan, regardless of regional linguistic variation.
Keywords: Automatic Speech Recognition, Kyrgyz Language, Dialect Adaptation, Low-Resource Languages, Parameter-Efficient Fine-Tuning, LoRA, Word Error Rate

Intelligent Agricultural Automation- Deep Learning Approach to Disease Detection and Irrigation Management

Emil Amrakulov[1], Aibike Moldobaeva[1], Bayastan Karimbaev[1], Seyitali İlyas[1]
[1] Kyrgyz-Turkish Manas University

There is a growing trend towards automation in the agricultural sector of Kyrgyzstan. This research proposes a new hybrid platform to increase crop yields through minimizing plant diseases by early detection of pathogenic symptoms on leaves and fruits and automating irrigation using available IoT systems. The research methodology includes the following stages: First, based on existing open sources, a database of samples of various diseases of agricultural crops cultivated in Kyrgyzstan was created. Second, a system for identifying pathogens was developed, based on the YOLOv8 deep learning model, which allows detecting visible signs of diseases on plant leaves and fruits. Third, an IoT microclimate system was integrated, providing automation of the irrigation process and remote control of it, which increases the convenience of farmers' work. To verify performance, comprehensive comparative experiments were conducted between the improved method and the traditional way of growing crops. The obtained results indicate the prospects of using a hybrid platform for optimizing agricultural production processes.
Keywords: Plant disease detection, IoT-based automation, Deep learning, Smart irrigation, Agricultural technology, YOLOv8

Küçük Ölçekli Seralar İçin Sensör Entegrasyonlu ve Yapay Zeka Tabanlı Akıllı Sulama Sistemi

Alibek Askarov[1], Ege Şensöz Karakaya[1], Naim Gassanov[1], Seyitali İlyas[1]
[1] Kyrgyz-Turkish Manas University

Bu çalışma, küçük ölçekli seralarda su israfını önlemek amacıyla toprakla ilgili parametreleri içeren hazır veriseti kullanarak ve derin öğrenme tabanlı görüntü işleme tekniklerini entegre eden düşük maliyetli bir yapay zeka sistemi sunmaktadır. Çalışma kapsamında;farklı bitki çeşitleri ve hastalık tespiti üzerine eğitilen çeşitli CNN(Evrişimli Sinir Ağları)karşılaştırılmalı olarak analiz edilmiştir.Elde edilen metrikler ışığında hesaplama maliyeti ve doğruluk oranı bakımından en yüksek performansı sergileyen mimari belirlenerek sistemin karar mekanizmasına dahil edilmiştir.
Anahtar Kelimeler: CNN,derin öğrenme, görüntü işleme, hastalık tespiti, sera otomasyonu, su tasarrufu.

Enhancing Systematic Literature Reviews: A Hybrid Framework Combining Robotic Process Automation and Generative AI

Medetkan Kutlu[1]
[1] Kyrgyz-Turkish Manas University

Systematic literature reviews (SLRs) are fundamental to evidence-based research but are traditionally characterized by labor-intensive manual processes that are both time-consuming and prone to human error. This study proposes a novel hybrid automation framework that integrates Robotic Process Automation (RPA) via UiPath with Generative Artificial Intelligence (GenAI) to streamline the end-to-end SLR workflow. The system automates keyword-based retrieval, metadata extraction, and semantic relevance screening, effectively bridging the gap between procedural automation and cognitive analysis. Experimental validation confirms that the proposed framework substantially reduces manual workload while preserving high data integrity and relevance quality. The results demonstrate that intelligent automation can significantly enhance both the speed and rigor of academic research, offering a scalable and interoperable solution for modern research workflows.

Assistive Software Application for Human Computer Interaction for with Disabilities

Begaiym Kudainazarova[1], Yalcin Albayrak[2]
[1] Akdeniz University
[2] Akdeniz Üniversitesi

Human–computer interaction (HCI) technologies are essential for improving digital accessibility and independent living for individuals with severe motor impairments, particularly those affected by neurological or neuromuscular disorders that substantially restrict voluntary movement. Because conventional input devices such as a mouse, keyboard, or touch interface are often inaccessible for these users, there is a critical need for alternative, low-cost, and practical interaction systems. In this study, a vision-based assistive HCI interface is proposed to enable hands-free computer control, communication, and basic environmental interaction. The developed system uses computer vision techniques to track head movement through the tip of the nose for cursor control, while intentional eye blinking is employed to perform click actions. A calibration stage defines the neutral reference position and active control boundaries, thereby improving interaction stability and usability. Beyond basic cursor movement, a dedicated accessible menu-based interface was developed to allow users to interact with commonly used applications such as YouTube, a web browser, and Telegram. The system also supports simple environmental control tasks, such as turning lights on and off, thereby extending its function beyond computer access into daily assistive living. In addition, an integrated help button enables users to send assistance requests via Telegram when needed. Compared with many existing assistive systems that rely on specialized hardware or provide limited functionality, the proposed approach offers a more accessible, multifunctional, and user-oriented solution. By integrating computer access, communication, emergency support, and environmental control into a unified interface, the system demonstrates strong potential for enhancing autonomy, inclusion, and quality of life for individuals with severe motor impairments.
Keywords: Human-Computer Interface, OpenCV, Motor Disabilities

Türk Dünyası Görsel Kültürünün Dijital Dönüşümü: Ortak Motiflerin Yapay Zekâ ile Yeniden Yorumlanması

Burak Erhan Tarlakazan[1]
[1] Grafik Tasarımı Bölümü (Güzel Sanatlar ve Tasarım Fakültesi), Kastamonu Üniversitesi

Türk dünyası; tarihsel, dilsel, kültürel ve coğrafi bağlar temelinde birbirleriyle ilişkili Türk dili topluluklarının oluşturduğu geniş bir kültürel alanı ifade etmektedir. Orta Asya'dan Anadolu'ya, Kafkasya'dan Balkanlar'a ve Sibirya'dan Orta Doğu'ya uzanan bu coğrafyada şekillenen ortak dil kökeni, mitolojik anlatılar, geleneksel sanat anlayışı ve motif repertuvarı; çok boyutlu bir kültürel bütünlük ve güçlü bir görsel kültür mirası ortaya koymaktadır. Bu bağlamda ortak motifler, kültürel kimliğin sürekliliğini sağlayan sembolik anlatım biçimleri olarak günümüzde dijital tasarım ve yapay zekâ uygulamaları aracılığıyla yeniden yorumlanmaya açık önemli bir üretim alanı sunmaktadır. Bu çalışma, Türk dünyasına ait geleneksel görsel kültür mirasının özellikle ortak motifler üzerinden yapay zekâ destekli tasarım süreçleriyle nasıl yeniden üretilebileceğini ve çağdaş grafik tasarım diline nasıl aktarılabileceğini incelemektedir. Halı, kilim, dokuma, mimari bezeme ve süsleme sanatlarında görülen motiflerin dijital ortama aktarılması, veri seti oluşturma süreçleri ve üretken yapay zekâ temelli görselleştirme teknikleri kapsamında deneysel uygulama önerileri geliştirilmektedir. Çalışma, ortak motiflerin yapay zekâ ile yeniden görselleştirilmesinin yalnızca teknik bir dönüşüm değil; aynı zamanda kültürel sürekliliğin dijital ortamda yeniden kurulmasına katkı sağlayan yaratıcı bir tasarım yaklaşımı sunduğunu ortaya koymayı amaçlamaktadır. Bu yönüyle araştırma, Türk dünyası görsel kültürünün dijital çağda sürdürülebilirliğine yönelik yeni üretim modellerini tartışmayı hedeflemektedir.
Anahtar Kelimeler: Türk dünyası, görsel kültür, motif, yapay zeka

Akıllı Şehir Altyapısı için Yapay Zeka Tabanlı Risk Bilinçli Varlık Yönetimi

Yılmaz Avcı[1], Cevat Özarpa[1]
[1] Ankara Medipol Üniversitesi

Bu çalışma, fiziksel varlıkların, dijital verilerin ve siber güvenliğin sıkı bir şekilde birbirine bağlı olduğu siber-fiziksel sistemlerin artan karmaşıklığına çözüm bulmak amacıyla, akıllı şehir altyapısı için yapay zekâ (YZ) tabanlı, risk bilincine sahip bir varlık yönetim çerçevesi önermektedir. Geleneksel varlık yönetimi ve güvenlik yaklaşımları genellikle birbirinden bağımsız olarak ele alınmakta ve bu da dinamik kentsel ortamlardaki etkinliklerini sınırlamaktadır. Bu sınırlamayı aşmak için, önerilen çerçeve, varlık yaşam döngüsü yönetimi ve bilgi güvenliği uygulamalarını, ISO 55001 ve ISO/IEC 27001 gibi uluslararası standartlarla uyumlu, YZ odaklı bir mimari içinde entegre etmektedir. Model, varlık, veri, güvenlik ve YZ bileşenleri de dahil olmak üzere birden fazla katmandan oluşmaktadır. Makine öğrenme teknikleri, IoT cihazlarından, coğrafi bilgi sistemlerinden (GIS) ve operasyonel kayıtlardan toplanan gerçek zamanlı veriler kullanılarak dinamik risk puanları oluşturmak için kullanılmaktadır. Bu, kentsel altyapı sistemleri için öngörücü bakım, anormallik tespiti ve gelişmiş durumsal farkındalık sağlamaktadır. Enerji dağıtım ağlarından esinlenerek yapılan vaka odaklı bir değerlendirme, önerilen yaklaşımın arıza tahmin doğruluğunu artırdığını, yanıt sürelerini azalttığını ve akıllı şehir ortamlarında yüksek riskli bölgelerin belirlenmesini desteklediğini göstermektedir. Sonuçlar, proaktif ve uyarlanabilir karar alma süreçlerini mümkün kılmak için yapay zekanın varlık yönetimi süreçlerine entegre edilmesinin etkinliğini vurgulamaktadır. Bu çalışmanın temel katkısı, varlık yönetimi ve siber güvenliği birleştiren, ölçeklenebilir ve pratik bir çözüm sunan, yapay zeka odaklı birleşik bir çerçeve geliştirmektir; bu çözüm, dayanıklı ve akıllı şehir altyapı sistemleri için idealdir.

Anahtar Kelimeler: Yapay Zeka, Risk Bilinçli Varlık Yönetimi, Akıllı Şehir Altyapısı, Siber-Fiziksel Sistemler, Tahmine Dayalı Bakım, Siber Güvenlik, Jeo-uzamsal Analiz

LLM-Based Security Assessment in Banking Payment Systems: Fraud Detection and API Vulnerability Analysis in Microservices Architecture

Taalai Almasova[1]
[1] Kyrgyz-Turkish Manas University

The rapid digitalization of banking payment systems has created security vulnerabilities at two levels: fraudulent transaction detection and exploitation of internally exposed API routes. Multi-partner payment ecosystems built on microservices architecture expose internal REST API endpoints — including payment initiation, balance inquiry, and transaction history routes — vulnerable to OWASP API Security Top 10 attacks. Traditional ML models such as XGBoost and Random Forest fail to protect these systems: they cannot reason semantically about API structure, operate as black boxes violating GDPR transparency requirements, and do not generalize across heterogeneous transaction formats from multiple partners. This study evaluates the effectiveness of LLM-based approaches for dual-layer security assessment in multi-partner banking payment systems: fraud detection at the transaction level and OWASP API Top 10 vulnerability identification in internally exposed payment microservice routes. A comparative experimental methodology is employed using the Kaggle Credit Card Fraud Detection dataset (284,807 transactions) and IEEE-CIS dataset (590,540 transactions), alongside real internally exposed payment API endpoints from an active banking microservices system. Baseline models include XGBoost and Random Forest; the LLM approach uses GPT-4 with markdown serialization. SHAP analysis addresses GDPR explainability requirements. No existing study simultaneously addresses fraud detection and OWASP API vulnerability assessment of internally exposed routes using LLMs in a microservices payment context. Results are expected to show competitive LLM fraud detection accuracy, detection of over 70% of OWASP vulnerabilities in real internal routes, and GDPR-compliant explanations through SHAP and Chain-of-Thought reasoning.

Keywords: large language models, fraud detection, API security, OWASP, banking microservices, internal route exposure, explainable AI, payment systems

Yapay Zeka Destekli Akıllı Şehir ve Akıllı Turizm Sistemleri: Türkiye'de Sürdürülebilir Turizm için Veri Odaklı SWOT Çerçevesi

Reşide Sağlam[1], Cevat Özarpa[2]
[1] Medipol Üniversitesi
[2] Ankara Medipol Üniversitesi

Dijital dönüşümün hızlı evrimi, kentsel yönetim ve turizm ekosistemlerini önemli ölçüde yeniden şekillendirerek, akıllı şehirleri ve akıllı turizmi sürdürülebilir kalkınmanın kritik itici güçleri olarak konumlandırmıştır. Bu çalışma, yapay zekâ (YZ) destekli bir SWOT analiz çerçevesi aracılığıyla Türkiye'deki akıllı şehir ve akıllı turizm uygulamalarını değerlendirmeyi amaçlamaktadır. Geleneksel nitel SWOT yaklaşımlarından farklı olarak, bu araştırma, turizm ve kentsel veri kaynaklarına dayalı olarak güçlü yönleri, zayıf yönleri, fırsatları ve tehditleri sistematik olarak sınıflandırmak için Doğal Dil İşleme (NLP) ve duygu analizi de dahil olmak üzere YZ destekli veri analizini entegre etmektedir. Çalışma, veri odaklı stratejik değerlendirme modeli oluşturmak için açık belediye veri kümelerinden, turizm platformlarından ve önceki akademik çalışmalardan elde edilen ikincil verileri kullanmaktadır. Bulgular, Türkiye'nin güçlü bir dijital altyapıya ve yüksek turizm potansiyeline sahip olduğunu; ancak veri parçalanması, siber güvenlik riskleri ve dengesiz teknolojik benimseme gibi zorlukların devam ettiğini ortaya koymaktadır. YZ entegrasyonu, gerçek zamanlı karar verme, tahmine dayalı turizm analizi ve kişiselleştirilmiş ziyaretçi deneyimleri için önemli bir kolaylaştırıcı olarak tanımlanmıştır. Çalışma, gelişmekte olan ekonomilerdeki akıllı turizm ekosistemlerine uyarlanmış yeni bir YZ destekli SWOT çerçevesi önererek literatüre katkıda bulunmaktadır. Ayrıca, yapay zekanın kentsel turizm yönetimine entegrasyonu, sürdürülebilirliğin artırılması ve rekabet avantajının iyileştirilmesi için politika önerileri sunmaktadır. Bulgular, akıllı destinasyon yönetiminde yapay zeka teknolojilerinden yararlanmak isteyen politika yapıcılar, şehir planlamacıları ve turizm paydaşları için değerli bilgiler sunmaktadır.
Anahtar Kelimeler: Akıllı Turizm, Akıllı Şehir, Yapay Zeka, SWOT Analizi, Dijital Dönüşüm, Sürdürülebilir Turizm

İş İlişkisi Bağlamında Yapay Zekânın Etik Kullanımı

İştar Urhanoğlu[1], Eda Manav Özdemir[2], Damla Kaynar Kiroğlu[1], Beyza Aydin[3]
[1] Hacettepe Üniversitesi
[2] Hâkim Ankara, Türkiye
[3] İzmir Katip Çelebi Üniversitesi Hukuk Fakültesi

Yapay zekâ sistemleri, günümüzde çevre, sağlık, finans, kamu ve istihdam dâhil olmak üzere pek çok alanda yoğun şekilde kullanılmaktadır. Bu sistemler, iş ilişkisinin kurulmasından sona ermesine kadar tüm aşamalarda işverenlerin karar alma süreçlerine dâhil edilmektedir. Özellikle iş alım sürecinde ilanların hazırlanması, adaylara ulaşılması, özgeçmişlerin analiz edilmesi ve mülakatların yürütülmesi gibi aşamalarda; iş ilişkisinin devamında ise performans değerlendirmesi, çalışma sürelerinin belirlenmesi, görev dağılımı ve ücretlendirme gibi hususlarda yaygın olarak kullanılmaktadır. Yapay zekâ sistemleri, verimlilik artışı sağlamakta; zaman ve maliyet tasarrufu sunmakta ve karar alma süreçlerinde tutarlılığı güçlendirmektedir. Bununla birlikte, işçilerin kişisel verilerin ihlali, "black box" özelliği nedeniyle şeffaflık ve açıklanabilirlik sorunları, algoritmik ayrımcılık, dijital dışlanma ve işçilerin temel haklarının ihlali gibi ciddi riskleri de beraberinde getirmektedir. Bu riskler karşısında kabul edilen AB Yapay Zekâ Tüzüğü, risk temelli bir yaklaşım benimsemekte ve iş ilişkilerinde kullanılan yapay zekâ sistemlerini "yüksek riskli" kategorisinde değerlendirmektedir. Bu kapsamda Tüzük'te işverenlere kapsamlı yükümlülükler getirilmiş; özellikle çalışanların duygusal durumlarının biyometrik veriler aracılığıyla analizine dayalı sistemler yasaklanmıştır. Türk hukukunda yapay zekâya ilişkin henüz doğrudan bir kanuni düzenleme bulunmamaktadır. Bununla birlikte düzenleyici çalışmalar hız kazanmış olup 2021-2025 Ulusal Yapay Zekâ Stratejisi hem küresel rekabet gücünün artırılmasını hem de yapay zekâ teknolojilerinin etik ve hukuki sınırlar içinde geliştirilmesini hedeflemektedir. Bu çalışma, literatür taraması yöntemiyle iş ilişkilerinde yapay zekânın etik kullanımını uluslararası düzenlemeler ile ülke uygulamaları çerçevesinde, incelemekte; Türk iş hukuku ile karşılaştırmalı olarak mevcut düzenlemelerin yeterliliğini ve hukuki boşlukları analizini amaçlamaktadır.

Anahtar Kelimeler: Yapay zekâ etiği, iş ilişkisi, algoritmik ayrımcılık, AB Yapay Zekâ Tüzüğü, kişisel verilerin korunması.

AI-Powered Cryptographic Agility for Quantum-Resilient Smart Grids: Challenges and Post-Quantum Solutions

Cevat Özarpa[1], Suat Arapoğlu[2], İsa Avcı[3]
[1] Ankara Medipol Üniversitesi
[2] Ankara Medipol University
[3] Karabük Üniversitesi

This study aims to analyze the risks of quantum computing technology on smart grid security protocols and evaluate the impact of NIST-standardized Post-Quantum Cryptography (PQC) algorithms on operational stability. In this context, AI-powered adaptive security mechanisms are explored as a dynamic solution to maintain system stability under varying network and threat conditions. The research encompasses a performance analysis of ML-KEM, ML-DSA, and SPHINCS+ algorithms across 10, 100, and 1,000-node Advanced Metering Infrastructure (AMI) scenarios, utilizing end-to-end latency, jitter, and packet delivery ratios as key performance indicators.

Since existing literature is largely limited to theoretical data or hardware prototypes, this study fills a critical gap by examining the impact of PQC-induced overheads on real-time control loops and cyber-physical system (CPS) stability in large-scale smart grids. Methodologically, a two-layer approach is adopted, combining algorithmic benchmarking via the liboqs framework with large-scale network simulations using GridLAB-D and OpenDSS platforms, all supported by AI-based analytical modeling.

Experimental results demonstrate that ML-KEM and ML-DSA meet sub-50 ms latency requirements, ensuring operational continuity; however, the high signature size of SPHINCS+ threatens critical stability margins. Furthermore, AI-based mechanisms are found to minimize performance degradation by dynamically switching between PQC schemes. By modeling cryptographic latency as a ”transport lag” (tau) in control systems and establishing the mathematical link between cybersecurity and physical stability, this research provides a strategic roadmap for IEC 62351 and IEEE 2030.5 standards. Ultimately, the study proposes an AI-powered cryptographic agility framework that balances cyber resilience with system stability for energy sector stakeholders.

Keywords: Post-Quantum Cryptography, Smart Grid Security, Control System Stability, ML-KEM, ML-DSA, SPHINCS+, Cyber-Physical Systems.

Çocuk Edebiyatında Dijitalleşme: Dijital Hikâye Anlatılarının ve Karakterlerin Okur Üzerindeki Etkisi

Müzeyyen İnalöz[1]
[1] Ankara Hacı Bayram Veli Üniversitesi

Çocuk edebiyatı, dijitalleşme süreciyle birlikte geleneksel anlatı biçimlerinden farklı olarak çoklu ortam özellikleriyle yeniden şekillenmektedir. Bu bağlamda dijital hikâye anlatıları, metin, görsel, ses ve animasyonu bir araya getirerek çocuk okura daha etkileşimli bir okuma deneyimi sunmaktadır. Bu çalışma, dijital hikâye anlatılarının ve karakterlerin çocuk okur üzerindeki etkisini incelemeyi amaçlamaktadır. Araştırmanın örneklemini TRT Çocuk Kitaplık platformunda yer alan Cemile Midilliye Biniyor, Mutlu Nasıl İyileşecek ve Kayıp Orman Şarkısı adlı dijital hikâyeler oluşturmaktadır. Bu kapsamda çalışma, nitel araştırma desenlerinden betimsel analiz yöntemine dayalı olarak yürütülmüş; belirlenen metinler anlatı yapısı, görsellik, animasyon kullanımı ve metin–görsel ilişkisi ölçütleri çerçevesinde incelenmiştir. İnceleme sonucunda, dijital hikâyelerde görsel ve hareketli unsurların anlatıyı destekleyerek okurun dikkatini çektiği ve anlamayı kolaylaştırdığı görülmüştür. Metin ile animasyon arasındaki eşzamanlılık, anlatının hem okunarak hem de izlenerek deneyimlenmesini mümkün kılmaktadır. Bununla birlikte, yoğun görsel-işitsel unsurların dikkat odağını metinden görsele kaydırma potansiyeli taşıdığı da belirlenmiştir. Bu çalışma, dijital çocuk edebiyatında dijital hikâye anlatılarının estetik ve yapısal özelliklerini ortaya koyması bakımından önem taşımaktadır. Elde edilen bulguların, dijital anlatıların çocukların okuma alışkanlıkları ve algı süreçleri üzerindeki etkilerine ilişkin literatüre katkı sağlaması öngörülmektedir.
Anahtar Kelimeler: Çocuk edebiyatı, dijitalleşme, dijital hikâye anlatımı, okur etkisi.

Yabancı Dil Olarak Türkçe Öğretiminde Sanal Gerçeklik (VR) Tabanlı Uygulama Tasarım Kriterleri

Ahmet Buğra İnalöz[1]
[1] Kırgızistan-Türkiye Manas Üniversitesi Yabancı Diller Yüksekokulu, Bişkek

Sanal gerçeklik teknolojisi, kullanıcıya üç boyutlu, etkileşimli ve duyusal açıdan zengin öğrenme ortamları sunarak hedef dilin gerçek yaşam bağlamında deneyimlenmesini mümkün kılmaktadır. Bu yönüyle VR, yabancı dil öğretiminde bağlamsal öğrenmeyi desteklemekte; öğrencilerin hedef dili ve kültürü doğal ortamlarda gözlemleme ve kullanma fırsatı elde etmesini sağlamaktadır. Yapılan araştırmalar, VR tabanlı dil öğretim uygulamalarının özellikle dinleme ve konuşma becerilerinin geliştirilmesinde etkili olduğunu ve dil-kültür bütünlüğünü desteklediğini ortaya koymaktadır. Bununla birlikte, VR uygulamalarının eğitimde etkili olabilmesi için pedagojik tasarım ilkeleri ile desteklenmesi gerekmektedir. Dil öğretiminde VR kullanımının dinleme, konuşma, yazma ve kelime öğretimine katkı sağladığı, ancak uygulamaların pedagojik yapı, kullanıcı deneyimi ve etkileşim düzeyi açısından farklılıklar gösterdiği belirtilmektedir. Bu nedenle VR tabanlı Türkçe öğretim uygulamalarının yalnızca teknolojik boyutlarıyla değil; pedagojik uygunluk, kültürel bağlam, etkileşim düzeyi, oyunlaştırma unsurları, geri bildirim mekanizmaları ve erişilebilirlik ilkeleri dikkate alınarak sistematik biçimde tasarlanması gerekmektedir. Bu çalışmanın amacı, yabancı dil olarak Türkçe öğretiminde sanal gerçeklik teknolojilerinin etkili kullanımına yönelik tasarım kriterlerini belirlemektir. Bu kapsamda mevcut VR tabanlı dil öğrenme uygulamaları incelenmiş ve pedagojik, teknolojik ve kullanıcı deneyimi boyutları betimsel analiz yöntemiyle değerlendirilmiştir. Elde edilen bulgular, etkili bir VR tabanlı Türkçe öğretim uygulamasının temel bileşenlerini ortaya koyarak uygulayıcılar ve geliştiriciler için yol gösterici bir çerçeve sunmaktadır.
Anahtar Kelimeler: Sanal gerçeklik, yabancı dil olarak Türkçe öğretimi, eğitim teknolojileri, dijital öğrenme, betimsel analiz.

Deneysel Isı Transferi Süreçlerinin Dijital Dönüşümü: Nanoakışkan ve Türbülatör Verilerinde Makine Öğrenmesi Uygulaması

Özlem H. Dağdelen[1], Gözde Katırcıoğlu İşgüç[1], Ahmet Özcan[1], M. Ali Ülkü[2]
[1] Kapadokya Üniversitesi Nevşehir, Türkiye
[2] Dalhousie Üniversitesi Halifax, Kanada

Bu çalışmada, farklı ağırlık oranlarına sahip nanopartikül-su karışımları ve çeşitli türbülatör geometrilerine ait ısı transfer performansının yapay zekâ tabanlı makine öğrenmesi yöntemleri ile tahmin edilmesi amaçlanmıştır. Deneysel çalışmaların yüksek maliyet,uzun süre ve yoğun iş gücü gerektirmesi, mühendislik uygulamalarında veri odaklı ve hızlı karar destek mekanizmalarının geliştirilmesiniönemli hale getirmektedir. Bu kapsamda, 341 gözlemden oluşan veri seti kullanılarak beş farklı hedef çıktı için Random Forest, XGBoost veLightGBM algoritmaları uygulanmış, model performansları R2, RMSE, MAE ve MAPE metrikleri ile değerlendirilmiştir. Tüm modellerde R2 değerlerinin 0,99 ve üzerinde seyretmesi, modellerin hedef çıktılardaki değişimi açıklama gücünün mükemmele yakın olduğunu ortaya koymaktadır. İstatistiksel olarak R2 katsayısının 1'e yaklaşması ile MAPE hata oranlarının bu denli düşük seviyelerde kalması, tahminlerin doğruluk derecesinin oldukça yüksek olduğunu ve uygulanan algoritmaların veri seti üzerinde üstün bir öğrenme performansı sergilediğini kanıtlamaktadır.Buna göre, en düşük hata oranları Random Forest modelinde (%0,24–%0,33), ardından XGBoost modelinde (%0,35–%0,47) elde edilirken,LightGBM modeli daha yüksek MAPE değerleriyle (%0,63–%0,86) görece daha düşük tahmin hassasiyeti göstermiştir. Elde edilen bulgular, yapay zekâ destekli tahminleme yaklaşımlarının deneysel mühendislik süreçlerinin dijital dönüşümünde etkili biçimde kullanılabileceğini, ısı transfer performansının yüksek doğrulukla öngörülebileceğini ve böylece zaman ile maliyet açısından önemli avantajlar sağlanabileceğini göstermektedir. Bu yönüyle çalışma, yapay zekânın mühendislik sistemlerinde uygulanabilirliğine ve deneysel süreçlerin akıllı modelleme teknikleriyle desteklenmesine katkı sunmaktadır.
Anahtar Kelimeler: Isı Transferi Artışı, Nanopartiküller, Türbülatör, Nusselt Sayısı, Makine Öğrenmesi.

Intelligent Technologies for Closed-Loop Supply Chain Management: Applications, Sustainability Impacts, and Future Research Directions

Sabahattin Çiftçi[1], Yunus Emre Karaman[1], Ahmet Özcan[1], Gözde Sunman[1], Kürşat O. Şahin[1], M. Ali Ülkü[2]
[1] Cappadocia University
[2] Dalhousie University

Closed-loop supply chains (CLSCs) integrate forward logistics with reverse flows such as reuse, remanufacturing, recycling, and disposal, enabling organizations to pursue sustainability beyond efficiency or cost objectives. Despite their strategic importance, CLSCs are characterized by uncertainty, fragmented information, and complex interdependencies among stakeholders, which limit the effectiveness of traditional decision-support approaches. Recent advances in artificial intelligence (AI) offer significant potential to address these challenges, yet sustainability-oriented deployment in CLSC contexts remains limited. In this paper, we propose a comprehensive framework for deploying intelligent technologies in CLSC management, grounded in the Quadruple Bottom Line (QBL) perspective, which emphasizes the economic, environmental, social, and cultural dimensions of sustainability. The framework conceptualizes CLSCs as adaptive, data-enabled systems in which AI supports decision-making across all QBL dimensions rather than optimizing economic outcomes alone. It maps AI techniques, including machine learning, optimization, digital twins, and autonomous decision agents, to key CLSC functions, such as demand forecasting with returns, inventory and production planning, reverse logistics network design, and real-time operational control. At the core of the framework are closed-loop feedback mechanisms that enable continuous learning from QBL metrics throughout the product life cycle. Practical application scenarios in manufacturing and service supply chains are discussed, highlighting data integration, transparency, and human-AI collaboration in advancing QBL-aligned objectives. We conclude by identifying future research directions, including multi-criteria AI models for QBL trade-offs, cultural and institutional constraints, governance and ethics, and alignment with circular economy transitions.

An Analysis of Graphic Design Products by Comparing Them with Digital Artificial Images Created Using Artificial Intelligence

Muhammet İnceağaç[1]
[1] Kastamonu University

The aim of this study is to examine original graphic design products in terms of graphic design by comparing them with digital artificial visuals created using artificial intelligence. In the study, logo designs included in corporate identity designs that constitute brand designs and packaging designs promoting the brand's products were comparatively analyzed in terms of originality and ethics, using visuals generated with artificial intelligence support. The research method employed document analysis, a qualitative research design. The data used in the study were interpreted descriptively. The data in the study was selected from the logo and packaging designs designed by the 4th-year students of the Department of Graphic Design at Kastamonu University, Faculty of Fine Arts and Design, as part of the 'Design Workshop V' course for the 2023-2024 and 2024-2025 academic year. The original design works included in the research are limited to 5 original logo designs and the product (packaging) designs for these logos, in order to serve as examples. The original designs in the research were used with permission obtained from the students for inclusion in the study. In order to compare original designs with those generated by artificial intelligence, five different artificial intelligence tools were used to produce visuals based on the original designs. These AI tools are mostly limited to the commonly preferred programs actively used in visual production: ChatGPT, Gemini, Claude, Ideogram, and Runway. As a result of the research, it can be stated that artificial intelligence tools lack original qualities in a design context and that these tools will pose ethical issues for original designs.
Anahtar Kelimeler: Graphic Design, Artificial Intelligence, Corporate Identity, Packaging Design, Brand and Product Design

AI-Assisted SCADA Systems in Water Treatment and Distribution in Smart Cities: Attacks, Prevention, and Risk Mitigation

Hadiya Daud Ismail[1], Cevat Ozarpa[2]
[1] Ankara Medipol University, BSc Computer Engineering, Integrated MSc Cybersecurity,
[2] Ankara Medipol university, Graduate School of Natural and Applied Sciences, Department of Cyber Security

Water treatment and distribution systems are critical components of smart city infrastructures, relying on SCADA systems to manage physical processes such as pumping, pressure regulation, and chemical dosing. With the increasing integration of IT, IIoT, and cloud-based services, these systems have evolved into complex cyber-physical environments, significantly expanding their attack surface. Legacy industrial communication protocols, including Modbus and DNP3, lack robust security mechanisms such as authentication and encryption, making SCADA systems vulnerable to cyber threats such as DoS, spoofing, and command injection attacks.This study presents a structured analytical review of cybersecurity risks in water SCADA systems, focusing on protocol vulnerabilities, cyber-attack patterns, detection techniques, and architectural weaknesses. The analysis reveals that existing solutions are largely fragmented, addressing isolated components rather than providing a comprehensive system-wide protection strategy. To address this gap, this paper proposes an AI-enhanced multi-layer cybersecurity framework tailored for water SCADA infrastructures in smart cities. The proposed model integrates protocol hardening, secure communication, network segmentation, and IIoT/cloud protection with AI-assisted capabilities, including anomaly detection, predictive analytics, threat intelligence, and decision support. Unlike traditional static security approaches, the framework introduces adaptive and context-aware defense mechanisms through cross-layer AI integration. The findings suggest that effective protection of water SCADA systems requires a unified, layered, and intelligent cybersecurity strategy that combines technical controls with operational and governance mechanisms. The proposed framework contributes to improving detection accuracy, reducing response time, and enabling proactive risk mitigation, thereby enhancing the resilience and sustainability of smart water infrastructures.
Keywords: SCADA, Water Infrastructure, Cybersecurity, Industrial Control System (ICS), IIoT

AI-Assisted Leakage Risk Prediction in Smart City Infrastructures: An Application on the İzmirgaz Natural Gas Network

Ahmet Yetik[1], Cevat Ozarpa[2], Ibrahim Horoz[1], Oykum Gulergul[1], Onur Kunduraci[1], Ender Sunerli[1]
[1] Izmir Dogalgaz Dagitim Ticaret ve Taahhut A.S.
[2] Ankara Medipol University

Early detection of gas leak risks in the natural gas distribution infrastructure in smart city environments is crucial for public safety, supply continuity, and efficiency. This study proposes an analytical model an AI-supported, data-driven analytical model to predict segments with elevated leak risk of the İzmirgaz natural gas distribution network. The model integrates three enterprise data sources: emergency call records (187), data from mobile gas leak detection devices, and key infrastructure data obtained from the Geographic Information System (GIS). The integrated dataset includes variables such as incident time, location, type of intervention, year of manufacture, manufacturer, pipe diameter and type, contractor-related attributes, and spatial characteristics of the infrastructure. By combining spatial and temporal data, the relationships between past leak patterns and infrastructure characteristics are analyzed to identify the key factors associated with leak occurrence.

A segment-level risk assessment approach is presented using machine learning methods such as logistic regression, decision trees, and random forests, based on the proposed framework. The proposed model supports the transition from reactive, event-driven maintenance to predictive and risk-based maintenance planning. Furthermore, risk assessments can be visualized in a GIS-based decision support tool, along with field intervention planning, maintenance prioritization, and field response planning.

At the intersection of artificial intelligence, big data analytics, and smart infrastructure management, this study proposes a scalable and applicable model for the proactive assessment of leak risks in natural gas distribution networks. The proposed method aims to increase efficiency, reduce maintenance costs, and contribute to a safer and more resilient infrastructure.

Keywords: Natural gas distribution, gas leak prediction, machine learning, GIS, big data analytics, decision support system, predictive maintenance, smart cities

Physics-Informed Machine Learning for Enhanced Modeling of α- Decay Systematics

Sihem Berbache[1], Serken Akkoyun[2,3], Sadiye M.Cakmak[4]
[1] Physics Department, Faculty of Science of Matter, University of Batna 1, Batna 05000, Algeria
[2] Department of Physics, Faculty of Sciences, Sivas Cumhuriyet University, Sivas, Türkiye
[3] School of Mathematics and Physics, University of Surrey, Guildford, Surrey, United Kingdom
[4] Osmangazi University Program of Medical Monitoring Techniques Eskisehir Türkiye

The α-decay half-life ($T_{1/2}$) and decay energy (Q_α) are governed by quantum tunneling through the Coulomb barrier and are classically described by the Geiger–Nuttall law. While this law captures the global linear dependence between $\log(T_{1/2})$ and Q_α?¹?², it fails to account for systematic deviations arising from nuclear structure effects, including shell closures, pairing correlations, deformation, and α-cluster preformation probability.In this work, we propose a physics-informed machine learning framework, termed PI-BiSE (Physics-Informed Boosted Stacking Ensemble), to model these deviations and enhance predictive accuracy for both Q_α and $\log(T_{1/2})$. The approach integrates multiple regression learners within a stacked ensemble architecture, combining nonlinear representation learning with robust statistical generalization. The model is trained on evaluated experimental nuclear data and incorporates physically motivated descriptors such as shell-distance parameters, isospin asymmetry, and pairing terms.The PI-BiSE model achieves high predictive performance, with coefficients of determination reaching $R^2 \approx 0.96$ for half-life prediction, indicating strong agreement with experimental observations. Beyond accuracy, the model captures nonlinear deviations from the Geiger–Nuttall law, effectively learning correction terms that encode underlying nuclear structure effects not explicitly described in conventional semi-empirical or mean-field models. These corrections can be interpreted as an emergent multivariate functional representation of nuclear structure contributions.These findings demonstrate that machine learning can serve as a perturbative extension to established physical laws, enabling a hybrid modeling paradigm in which data-driven corrections complement theoretical formulations. The proposed framework provides a robust and generalizable tool for investigating α-decay systematics across the nuclear chart, including regions with sparse experimental data.
Keywords: α-decay, Geiger–Nuttall law, physics-informed machine learning, stacking ensemble, nuclear structure, half-life prediction.

Deepfake Technology: On The Brink of A Digital Identity Crisis

Muhammed Seyrek[1], Merve Yıldırım[2]
[1] Erzurum Teknik Üniversitesi
[2] Karadeniz Teknik Üniversitesi

This paper examines deepfake technology — AI-generated synthetic media — covering its underlying mechanisms, societal and individual threats, and current detection and mitigation strategies. Deepfakes are primarily created using Generative Adversarial Networks (GANs), which can produce highly realistic fabricated images, audio, and video. As the technology becomes increasingly accessible, it poses serious risks including political manipulation, financial fraud, and personal reputation damage. The paper also discusses AI-based detection tools, emerging legal frameworks, and the critical role of media literacy in combating this threat. The discussion is situated within the broader context of AI governance and digital transformation, themes that are central to the TURKWAI 2026 congress agenda.
Keywords: Deepfake, Artificial Intelligence, GAN, Cybersecurity, Disinformation, Digital Identity, Synthetic Media

PI Denetleyici Parametrelerinin Pekiştirmeli Öğrenme Tabanlı Otonom Optimizasyonu

Muhammet Ali Bulut[1], Aykut Kaya[1], Seyitali İlyas[2], İlhan Tunç[1]
[1] Bursa Teknik Üniversitesi
[2] Kyrgyz-Turkish Manas University

Endüstriyel kontrol sistemlerinde PI denetleyiciler yaygın olarak kullanılmakla birlikte, optimum parametre ayarı zaman alıcı bir süreçtir ve dinamik çalışma koşullarında daha da zorlaşmaktadır. Bu çalışmada, bir sistemin PI parametrelerini otonom olarak optimize etmek amacıyla pekiştirmeli öğrenme tabanlı bir ajan geliştirilmiştir. Önerilen yöntemde, sistem performansı zaman ağırlıklı mutlak hata integrali ve aşım cezasını içeren bir maliyet fonksiyonu ile değerlendirilmiştir. Geliştirilen pekiştirmeli öğrenme temelli ajanın durum uzayı maliyet ve aşım seviyelerine göre tanımlanmış; eylem uzayı ise PI kazanç parametrelerinin artırılması, azaltılması veya sabit tutulması şeklinde oluşturulmuştur. Ajanın eğitim sürecinde keşif ve sömürü stratejileri birlikte kullanılmıştır. Ayrıca, belirlenen performans kriterine ulaşma süresini azaltmaya yönelik bir ödül-ceza mekanizması oluşturulmuş ve maliyet değerine bağlı olarak epizodun erken sonlandırılmasına imkân veren bir yapı tasarlanmıştır. Simülasyon sonuçları, eğitilen ajanın referans değere düşük aşım ve kısa yerleşme süresi ile ulaştığını ve PI parametre optimizasyonunda başarılı bir performans sergilediğini göstermektedir. Elde edilen bulgular, pekiştirmeli öğrenme tabanlı yaklaşımın insan müdahalesini azaltarak endüstriyel kontrol uygulamaları için uyarlanabilir ve etkili bir çözüm sunduğunu ortaya koymaktadır.
Anahtar Kelimeler: Pekiştirmeli öğrenme, adaptif pi kontrol, kontrol sistemleri

Uzun Mesafeli Elektrikli Araç Yolculukları İçin Pekiştirmeli Öğrenme Tabanlı Şarj Planlama Asistanı

Emre Yalçın[1], Saadettin Yıldırım[1], Seyitali İlyas[2], İlhan Tunç[1]
[1] Bursa Teknik Üniversitesi
[2] Kyrgyz-Turkish Manas University

Elektrikli araçlar, sundukları çeşitli avantajlar ve çevre dostu yapıları nedeniyle giderek daha fazla tercih edilmektedir. Bununla birlikte, kullanıcılar özellikle uzun mesafeli yolculuklarda şarj planlamasında güçlük yaşayabilmektedir. Bu durum, etkin şarj planlamasını önemli bir araştırma problemi haline getirmektedir. Bu problem, dinamik ve çok değişkenli bir karar verme süreci niteliği taşımaktadır. Araç dinamikleri, şarj altyapısı, trafik, çevresel koşullar ve sürücü davranışı gibi birçok unsurun eşzamanlı etkisi altında şekillenmektedir. Bu çalışmada, fiziksel çevre koşulları ve sürücü davranışını birlikte dikkate alarak karar verebilen pekiştirmeli öğrenme temelli bir şarj planlama asistanı geliştirilmiştir Çalışmada kullanılan simülasyon ortamı, Google Directions API ile Open Charge Map kaynaklarından elde edilen veriler kullanılarak oluşturulmuştur. Önerilen yöntemde, batarya seviyesi, konum, yol eğimi, trafik, hava koşulları, şarj istasyonu özellikleri ile sürücüye bağlı bekleme toleransı, mola ihtiyacı ve menzil kaygısı gibi çok sayıda parametre, pekiştirmeli öğrenme ajanının karar verme sürecine dahil edilmiştir. Ayrıca, yaklaşımın farklı araç tiplerine ve sürücü profillerine uyarlanabilir olması, akıllı elektrikli araç navigasyon sistemleri açısından uygulanabilir bir alternatif sunduğunu göstermektedir.
Anahtar Kelimeler: Elektrikli Araçlar, Pekiştirmeli Öğrenme, DQN, Şarj planlaması

Türk Dünyası Başkentlerinde Dijital Yönetişim ve Akıllı Şehir Uygulamaları: Ankara, Bişkek, Astana, Bakü ve Taşkent'in Karşılaştırmalı Analizi (2021–2026)

Nefise Uzun[1], Cevat Özarpa[1]
[1] Ankara Medipol University

Türk Dünyası başkentlerinde dijital dönüşüm stratejik bir ivme kazansa da yönetişim boyutu ve teknoloji entegrasyonuna dair bütüncül çalışmalar sınırlıdır. Bu çalışma; Ankara, Bişkek, Astana, Bakü ve Taşkent'in 2021–2026 dönemi dijital yönetişim ve akıllı şehir politikalarını karşılaştırmalı incelemektedir. Analiz; e-devlet olgunluğu, akıllı altyapı, siber güvenlik ve yapay zekâ tabanlı veri analitiği boyutları üzerinden kurgulanmıştır. Çalışmada bazı göstergeler doğrudan başkent ölçeğinde, bazıları ise ulusal dijital yönetişim göstergeleri üzerinden değerlendirilmiştir. Bu nedenle bulgular, başkent uygulamaları ile ulusal dijital dönüşüm çerçevelerinin birlikte okunmasına dayalı nitel karşılaştırmalı bir analiz niteliğindedir. Araştırma, Türk Dünyası başkentlerinde dijital yönetişim ve akıllı şehir uygulamalarının farklı olgunluk düzeylerinde geliştiğini göstermektedir. Ankara güçlü e-devlet altyapısı ve katılımcı belediye uygulamalarıyla; Astana ileri dijital altyapı, e-devlet performansı ve yapay zekâ odaklı kamu politikalarıyla; Bakü ASAN/myGov deneyimi ve siber güvenlik kapasitesiyle; Taşkent hızlı dijitalleşme stratejisi ve e-devlet yatırımlarıyla; Bişkek ise dijital kamu reformu, devlet bulutu ve kapsayıcı bağlantı yatırımlarıyla öne çıkmaktadır. Bununla birlikte şehirler arasında veri yönetişimi, açık veri standardizasyonu, siber güvenlik olgunluğu, dijital okuryazarlık, kırsal-kentsel erişim farkı ve yerel düzeyde ölçülebilir performans göstergeleri bakımından iyileştirme alanları bulunmaktadır. Çalışma, bu saha deneyimlerinden hareketle Türk Devletleri Teşkilatı (TDT) çerçevesinde uygulanabilecek dört aşamalı bir "Dijital Yönetişim Yakınsama Modeli" önermektedir. Bu yönüyle bildiri, akademik literatüre katkı sunmakta ve bölgesel dijital iş birliği için uygulanabilir bir çerçeve ortaya koymaktadır.
Anahtar Kelimeler: dijital yönetişim, akıllı şehirler, Türk Dünyası, dijital dönüşüm, siber güvenlik, yapay zekâ.

Büyük Dil Modellerinin Türk Dünyasına İlişkin Olgusal Sorulardaki Halüsinasyon Eğilimlerinin Karşılaştırılması

Melike Çiftçi[1]
[1] Kapadokya Üniversitesi

Büyük dil modelleri (Large Language Models, LLM), pek çok alanda güçlü sonuçlar üretmesine karşın zaman zaman gerçekle örtüşmeyen ya da tamamen uydurulmuş bilgiler üretebilmektedir. "Halüsinasyon" olarak adlandırılan bu durum, modellerin eğitim verilerinde yeterince yer bulmayan konularda daha sık gözlemlenmektedir. Türk dünyasına ait tarihsel, kültürel ve coğrafi bilgiler bu açıdan özellikle kırılgan bir alan oluşturmaktadır; ancak bu bilgiler İngilizce merkezli eğitim verilerinde oldukça sınırlı biçimde temsil edilmektedir. Bu çalışmada, yaygın olarak kullanılan üç büyük dil modeli olan ChatGPT, Gemini ve Claude'un Türk dünyasına ilişkin Türkçe olgusal sorulardaki halüsinasyon eğilimleri karşılaştırmalı olarak incelenmiştir. Bu amaçla ortak Türk tarihi, Türk devletlerinin coğrafyası, Türk kültürü ve Türk dilleri gibi kategorileri kapsayan özgün bir soru veri seti oluşturulmuştur. Her üç model aynı sorularla test edilmiş; verilen yanıtlar güvenilir akademik ve ansiklopedik kaynaklarla karşılaştırılarak olgusal yanlışlık, bilgi uydurma ve belirsizlik türleri bazında sınıflandırılmıştır. Elde edilen bulgular hem model bazında hem de kategori bazında analiz edilerek modellerin güçlü ve zayıf yönleri ortaya konmuştur. Bu çalışma, Türk dünyası bilgisi özelinde LLM güvenilirliğini sistematik biçimde ele alan ilk çalışmalardan biri olup bu alandaki boşluğu doldurmayı hedeflemektedir.
Anahtar Kelimeler: Büyük dil modelleri, halüsinasyon, Türk dünyası, olgusal doğruluk, doğal dil işleme

Kırgızistan'da İş İlanlari ve Talep Edilen Becerilerin Analizi: JobScope KG Platformu

Ainuska Kaiymova[1], Gulshat Muhametjanova[1]
[1] Kyrgyz-Turkish Manas University

Bu çalışmada, Kırgızistan iş gücü piyasasındaki iş ilanlarını ve talep edilen becerileri analiz etmeye yönelik bir web platformu geliştirilmiştir. Platform, çevrimiçi iş ilanı verilerini otomatik olarak toplayarak doğal dil işleme ve makine öğrenmesi yöntemleriyle işlemektedir. Çalışmada özellikle mesleki beceri talebinin sektörel dağılımı, ücret düzeyleri ve iş ilanlarının anlamsal kümelenmesi üzerine odaklanılmıştır. Geliştirilen sistemde, iş ilanlarının vektörleştirilmesi ve kümelenmesi için TF-IDF ve K-Means algoritmaları, ücret tahmini için ise Random Forest yöntemi kullanılmıştır. Geliştirilen platform, büyük dil modeli tabanlı bir kariyer danışmanı işlevi de sunmaktadır. Bu danışman, gerçek piyasa verileriyle beslenerek kullanıcılara kariyer planlaması konusunda kişiselleştirilmiş öneriler sağlamaktadır. Önerilen yaklaşım; iş arayanlar, işverenler için veri odaklı karar destek aracı olarak kullanılabilir. Bu çalışma, yapay zekanın Türk dünyasında emek piyasası analizine uygulanmasına somut bir örnek sunmakta ve bölgesel dijital dönüşüme katkı sağlamayı hedeflemektedir.
Anahtar Kelimeler: iş gücü piyasası analizi, NLP, makine öğrenmesi, kariyer danışmanı, dijital dönüşüm, Kırgızistan

Медициналык веб-порталда AI-ассистентти интеграциялоонун моделдери

Alina Erlanova[1], Ella Abylaeva[1]
[1] Kyrgyz-Turkish Manas University

Саламаттык сактоо тармагындагы санарип трансформация пациенттердин медициналык кызматтарга жетүүсүн жеңилдетүүнүн негизги жолу болуп саналат. Бирок пациенттердин керектүү адисти тандоодо кыйналышы медициналык жардамга жетүүдөгү практикалык тоскоолдук бойдон калууда. Бул иш аталган көйгөйгө жооп катары жасалма интеллект ассистенти интеграцияланган көп ролдуу медициналык веб-порталды иштеп чыгууну максат кылат. Портал Django (Python) фреймворкунда иштелип чыккан жана Пациент, Дарыгер жана Клиниканын Администратору деген үч ролду колдойт. Системанын негизги өзгөчөлүгү — ЖИ ассистенти: пациент симптомдорун жазат, ал эми ассистент аларды талдап, тиешелүү медициналык адисти сунуштайт. Бул механизм адисти тандоо процессин жөнөкөйлөтүп, колдонуучунун убактысын кыскартат. Система ошондой эле онлайн жазылуу, медициналык карталар, клиникалар каталогу жана пикирлер модулдарын камтыйт. Интерфейс AJAX технологиясы аркылуу динамикалык жаңыртылат, ал эми ролго негизделген мүмкүндүктөрдү башкаруу коопсуздукту камсыздайт. Сунушталган чечим жасалма интеллектти медициналык маалыматтык системаларга ийгиликтүү интеграциялоо мүмкүн экендигин далилдеп, саламаттык сактоо кызматтарын санариптештирүүгө практикалык салым кошот.

Ачкыч сөздөр: жасалма интеллект, медициналык портал, Django, ЖИ ассистенти, санарип трансформация, пациент навигациясы, веб-иштеп чыгуу

Gamedev чөйрөсүндө жасалма интеллектти курал катары колдонуу: заманбап ыкмалар жана өнүгүү багыттары

Begimay Gulamkadyrova [1], Ella Abylaeva [1]
[1] Kyrgyz-Turkish Manas University

Оюн индустриясы генеративдик жасалма интеллекттин (GenAI) таасири астында башынан чоң өзгөрүүлөрдү көрүп жатат. Google Cloud (2025) маалыматы боюнча, оюн иштеп чыгуучулардын 90% ЖИ өз жумуш процесстерине толук киргизген, ал эми 97% анын индустрияны түп-тамырынан өзгөртүп жатканына ишенет. Бүгүнкү күндө генеративдик жасалма интеллект визуалдык ассеттерди түзүүдө, код жазууда, каармандардын диалогдорун генерациялоодо жана оюн дүйнөлөрүн процедуралык түзүүдө активдүү колдонулуп жатат. Midjourney, Stable Diffusion жана GitHub Copilot сыяктуу куралдар азыркы оюн иштеп чыгуучунун жумуш процессинин кадимки бөлүгүнө айланды. Бул өзгөчө кичинекей командалар үчүн маанилүү — ЖИ аларга ири студиялар менен атаандашууга жакшы мүмкүнчүлүк берет. Бирок технологияны көзөмөлсүз пайдалануу төмөн сапаттагы оюндардын агымын жаратып, авторлук укуктун чегин жоюп жатат. Иштеп чыгуучулар ошондой эле чыгармачылык процесстеги адамдык уникалдуулуктун жоголушунан коркушат. ЖИ-контентке байланышкан укуктук маселелер азырынча чечилбей жатат. Ошого карабастан, аң-сезимдүү колдонгондо ЖИ адамды алмаштыруучу эмес, анын чыгармачыл өнөктөшү болуп саналат. Геймдевдин келечеги технология менен адамдык чыгармачылыктын ортосундагы балансда.

Ачкыч сөздөр: Game development, GenAI, визуалдык ассеттер, код генерациясы, LLM, Stable Diffusion, чыгармачылык.

Турак Жай Бааларын Болжолдоо Үчүн Машиналык Окутуу Ыкмаларына Негизделген Интеллектуалдык Система

Турдуайым Канатбекова[1], Элла Абылаева[1]
[1] Кыргыз-Турк Манас Университети

Бул макалада машиналык окутуунун заманбап ыкмаларын колдонуу аркылуу турак жай баасын болжолдоо маселеси каралат. Изилдөөнүн актуалдуулугу кыймылсыз мүлк рыногунун өзгөрмөлүүлүгү жана турак жайдын баасын так аныктоого болгон муктаждык менен түшүндүрүлөт. Иштин алкагында маалыматтарды чогултуу жана алдын ала иштетүү жүргүзүлдү. Random Forest, Gradient Boosting жана XGBoost сыяктуу регрессиялык моделдер иштелип чыгып, салыштырылды. Гиперпараметрлерди тандоо кросс-валидация ыкмасы аркылуу аткарылды. Жыйынтыгында эң натыйжалуу модель тандалып алынып, колдонуучуга ыңгайлуу интерфейси бар программага киргизилди. Иштелип чыккан система кыймылсыз мүлк баасын практикалык баалоодо колдонулушу мүмкүн.

Ачкыч сөздөр: машиналык окутуу, турак жай баасы, болжолдоо, XGBoost, регрессия, маалыматтарды талдоо

Adana İlindeki Kavşak Tasarım Uygulamalarının Mikrosimülasyon Destekli Optimizasyonu

Gülşah Doğan[1]
[1] Adana Büyükşehir Belediyesi

Bu çalışma, kentsel ulaşım sistemlerinde kavşak tasarımının trafik işletimi, güvenlik ve sürdürülebilirlik üzerindeki etkilerini ortaya koymak amacıyla gerçekleştirilmiştir. Çalışma kapsamında, Adana ili Seyhan ilçesinde yer alan iki önemli kavşakta uygulanan tasarım ve düzenleme örnekleri incelenmiştir. İlk olarak, mevcut hemzemin kavşak düzeni ile üst geçitli Single Point Urban Interchange (SPUI) tasarımı, PTV VISSIM mikrosimülasyon yazılımı kullanılarak karşılaştırmalı biçimde modellenmiş ve performans açısından değerlendirilmiştir. İkinci olarak, Ali Bozdoğanoğlu Bulvarı ile Aliya İzzet Begoviç Bulvarı kesişiminde yer alan dönel adalı kavşakta gözlenen trafik tıkanıklığı ve güvenlik sorunlarının giderilmesine yönelik gerçekleştirilen fiziksel düzenleme ele alınmıştır. Analiz sonuçları, SPUI tasarımının kavşak hizmet düzeyini LOS F'den LOS D'ye yükselttiğini, ortalama gecikme süresini 166 saniyeden 54 saniyeye düşürdüğünü, ortalama araç hızını 10 km/saatten 25 km/saate çıkardığını ve yakıt tüketiminde %44 oranında iyileşme sağladığını ortaya koymuştur. Ayrıca, dönel adalı kavşakta yapılan fiziksel düzenleme sonrasında kaza verilerinde belirgin bir azalma olduğu belirlenmiştir. Elde edilen bulgular, mikrosimülasyon modellemesinin kentsel kavşak tasarım ve iyileştirme süreçlerinde karar destek aracı olarak önemli bir işleve sahip olduğunu; ekonomik, operasyonel ve güvenlik açısından etkin sonuçlar sunduğunu göstermektedir.
Anahtar Kelimeler: kavşak tasarımı, mikrosimülasyon, VISSIM, SPUI, hizmet düzeyi, trafik güvenliği, Adana

Siber Güvenlik Eğitiminde Üretici Yapay Zeka Destekli Otomatik Müfredat Tasarımı: NIST NICE ve MITRE ATT&CK Entegrasyonlu RAG ile Bloom Taksonomisi Tabanlı Bir Yaklaşım

Mehmet Sarıgil[1], Cevat Özarpa[1]
[1] Ankara Medipol University

Sürekli artan siber güvenlik tehditleri, siber güvenlik eğitim müfredatının da hızlıca güncellenmesini gerektirmektedir. Geleneksel eğitim müfredatı hazırlama yöntemleri, statik yaklaşım sergilemeleri ve alan uzmanına bağımlı olmaları nedeniyle bu ihtiyacı karşılayamamaktadır. Literatürde, yapay zeka destekli müfredat üretilmesi konusunda yayınlar olmasına rağmen, NIST NICE ve MITRE ATT&CK çerçevelerini, RAG pipeline'ında bütünleştirerek Bloom taksonomisiyle uyumlu modüler müfredat oluşturulması ve karşılaştırmalı değerlendirme yapılması yeterince ele alınmamaktadır. Bu çalışmada, siber güvenlik eğitim müfredatının hazırlanmasının statik yaklaşımdan ve uzman bağımlılığından kurtarılması, yapay zeka destekli olarak kısa sürede hazırlanması amacıyla kullanılabilecek bir model önerilmektedir. Modelde; NIST NICE, MITRE ATT&CK verileri ve üniversitelerin siber güvenlik eğitim müfredatları bilgi katmanı olarak kullanılmakta, elde edilen bilgiler vektörleştirilerek vektör veri tabanında depolanmaktadır. Üretim katmanında LLM modeli, RAG ile sorgulanan bağlamı kullanarak Bloom taksonomisinin kapsadığı düzeylere karşılık gelen öğrenme hedefleri üretmektedir. Çıktı olarak haftalık konu planı, öğrenme hedefleri ve değerlendirme stratejilerini içeren modüler ders planı oluşturulmaktadır. Çalışma kapsamında, modelin NIST NICE ve MITRE ATT&CK gibi verilerin güncellenmesine bağlı olarak müfredatı otomatik güncellendiği ve geleneksel yaklaşımlar ile elde edilemeyen kurumsal çevikliğe ulaşılabildiği görülmektedir. Model çıktısında, farklı programlar ve yetkinlik düzeyleri arasında müfredat tutarlılığı otomatik olarak sağlanmaktadır.
Anahtar Kelimeler: Eğitimde Dijital Dönüşüm, Siber Güvenlik Eğitimi, LLM, RAG, NIST NICE, MITRE ATT&CK, Bloom Taksonomisi

Cyberattacks on Biomedical Devices and AI-Based Detection

Ela Dizeci[1], Cevat Özarpa[2]
[1] Department of Biomedical Engineering, Middle East Technical University Ankara, Türkiye
[2] Department of Engineering and Natural Sciences, Ankara Medipol Üniversitesi, Ankara, Türkiye

The rapid growth of connected biomedical devices within the Internet of Medical Things has significantly improved healthcare services through real-time monitoring, remote diagnosis, and data-driven clinical decision-making. However, this connectivity also increases cybersecurity risks for devices such as pacemakers, insulin pumps, wearable sensors, patient monitors, and diagnostic systems. Cyberattacks including unauthorized access, ransomware, denial-of-service attacks, false data injection, firmware tampering, and manipulation of device communication can directly affect patient safety, data integrity, treatment accuracy, and healthcare service continuity.

This study focuses on the use of artificial intelligence-based detection methods against cyberattacks targeting biomedical devices. The proposed approach aims to identify abnormal behaviors by analyzing network traffic, sensor data, device communication patterns, operation timing, and system-level indicators. Machine learning and deep learning techniques such as Support Vector Machines, Random Forest, clustering algorithms, autoencoders, Convolutional Neural Networks, Recurrent Neural Networks, and Long Short-Term Memory networks are evaluated as possible detection mechanisms. In addition, the study discusses key challenges such as limited device resources, lack of standard datasets, false positives, explainability problems, and adversarial attacks against AI models. The study highlights the need for lightweight, explainable, adaptive, and clinically reliable AI-based intrusion detection systems for secure biomedical environments.

Keywords: Biomedical Devices; Cybersecurity, Internet of Medical Things; Artificial Intelligence; Intrusion Detection; Anomaly Detection; Explainable AI

An Assessment of Digital Ethics and Artificial Intelligence: The Case of Türkiye and the European Union

Mert Soğukoluk[1], Sibel Duygulu[1], Yeliz Bulut[1]
[1] Yozgat Bozok University

The rapid spread of digitalization and artificial intelligence across economic, administrative, and social spheres necessitates a reevaluation of ethical principles and rules in the digital environment. In particular, with the rise of artificial intelligence applications, the use of big data, the growing influence of online platforms, and the widespread adoption of algorithmic decision-making processes, artificial intelligence applications and the concept of digital ethics have become the focal point of current debates. This study aims to examine current developments and debates in the fields of digital ethics and artificial intelligence through a comparative lens, focusing on the European Union and Türkiye. Digital ethics and artificial intelligence are addressed within the framework of the protection of individuals' fundamental rights and freedoms, accountability, transparency, the security of personal data, and principles of justice. The research compares the European Union's normative and institutional approach to digital ethics with the current legal and administrative regulations in Türkiye. In this context, while evaluating the General Data Protection Regulation, the Digital Services Act, the European Union Declaration on Digital Rights and Principles for the Digital Decade, and ethical guidelines on artificial intelligence which form the foundation of the European Union's digital ethics policies, Türkiye's Law on the Protection of Personal Data, the Turkish Grand National Assembly's Artificial Intelligence Bill, and related secondary legislation have been examined. The study is based on document analysis, a qualitative research method. The findings reveal that the European Union has adopted a comprehensive, binding, and human rights-based governance model in the field of digital ethics, and that the AI Act it has established plays a regulatory role. In Türkiye, however, regulations related to digital ethics are primarily shaped within the framework of data security and legal compliance, and it is observed that ethical principles have not been sufficiently integrated into practice. Nevertheless, it is noted that awareness of digital ethics is gradually increasing in Türkiye and that the process of harmonization with the EU serves as a significant driving force in this area. It is noted that efforts related to artificial intelligence in Türkiye are gaining momentum and are increasingly focused on legal regulations in the field of artificial intelligence. The study aims to provide academic and practical contributions toward the development of ethics-based digital policies by highlighting the structural and managerial limitations Türkiye faces in the fields of digital ethics and artificial intelligence, as well as the EU's areas of application.
Keywords: Digital Ethics, Artificial Intelligence, The European Union, Türkiye

Endüstri 4.0 Kapsamında ERP ve MES Sistemlerinin Yapay Zeka Destekli Uyarlanması: Fabrika Göstergeleri Tabanlı Bir Yaklaşım

Abdulkadir Özarpa[1], Cevat Özarpa[1]
[1] Ankara Medipol Üniversitesi

Endüstri 4.0 ile işletmelerde ERP (Kurumsal Kaynak Planlama) ve MES (Üretim Yürütme Sistemleri) kullanımı hızla yaygınlaşmış olsa da bu sistemlerin fabrikalara etkin şekilde uyarlanması önemli bir sorun olarak varlığını sürdürmektedir. Mevcut çalışmalar çoğunlukla sistemlerin teknik özellikleri ve faydalarına odaklanırken, her fabrikanın kendine özgü üretim yapısı, organizasyonel kültürü ve dijital olgunluk seviyesine uygun uyarlama süreçleri yeterince ele alınmamaktadır. Bu çalışmada, ERP ve MES sistemlerinin uygulanmasında karşılaşılan bu uyarlama problemini çözmek amacıyla, yapay zeka destekli ve fabrika göstergelerine dayalı bir yaklaşım önerilmektedir.

Önerilen modelde, üretim hacmi, ürün çeşitliliği, iş gücü yapısı, süreç karmaşıklığı ve dijital olgunluk gibi temel fabrika göstergeleri dikkate alınarak mevcut durum analizi yapılmakta ve işletmeye özgü sistem gereksinimleri belirlenmektedir. Yapay zeka tabanlı karar destek mekanizması, bu göstergeleri analiz ederek her fabrika için en uygun ERP ve MES konfigürasyonunu önermektedir. Böylece standart çözümler yerine, işletmeye özel uyarlama stratejileri geliştirilmesi mümkün hale gelmektedir.

Çalışma kapsamında farklı özelliklere sahip üretim senaryoları üzerinden modelin uygulanabilirliği değerlendirilmiş ve göstergelere dayalı uyarlama yaklaşımının, sistemlerin etkinliğini ve uygulama başarısını artırdığı görülmüştür. Sonuç olarak, ölçülebilir göstergelere dayanmayan üretim sistemlerinin sürdürülebilir olmadığı, yapay zeka destekli uyarlama modellerinin ise dijital dönüşüm süreçlerinde kritik bir rol oynadığı ortaya konmuştur.

Anahtar Kelimeler: Endüstri 4.0, ERP, MES, Yapay Zeka, Dijital Dönüşüm, Üretim Yönetimi, Karar Destek Sistemleri

From Prompts to Vulnerabilities: Security Risks in AI-Generated IoT Code

Ayşe Rüveyda Bilik[1], Ayşe Rüveyda Bilik[2], Fadi Yılmaz[3]
[1] Cappadocia University
[2] Ankara Yıldırım Beyazıt University
[3] Qatar University

The increasing use of AI-assisted coding tools in Internet of Things (IoT) development has significantly lowered the barrier to entry, enabling even inexperienced developers to produce functional firmware. However, this convenience introduces a critical security risk: AI-generated code, particularly when guided by users with limited security awareness, may systematically embed vulnerabilities into IoT systems. This study investigates the nature and patterns of such vulnerabilities by generating a controlled corpus of AI-produced IoT firmware. To achieve this, the corpus is constructed using prompts that an inexperienced, security-unaware developer might produce, such as: "Develop code that copies the username into the user profile page after login." The corpus is generated using Google Gemini and spans various programming tasks written in Python, while deliberately incorporating insecure development practices.
The generated firmware is analyzed using dynamic fuzz testing, guided by the CWE Top 25 Most Dangerous Software Weaknesses. Beyond identifying vulnerabilities, the study focuses on extracting recurring "vulnerability signatures" that reveal how AI-generated code introduces specific security flaws. Finally, the study labels these vulnerabilities based on the patterns they exhibit in order to identify the attack surface of the AI-generated IoT firmware ecosystem.
Keywords: Artificial Intelligence, Internet of Things (IoT), AI-assisted Coding, Vulnerability Analysis, Fuzz Testing, Software Security, Vulnerability Signatures

Artificial Intelligence-Driven Polymeric Nanocarriers for Targeted Cancer Therapy: Design, Optimization and Future Perspectives

Deniz Yıldırım[1], Şevval Sena Bulut[1], Tarık Emre Taşkan[1], Cevat Özarpa[1]
[1] Department of Biomedical Device Technologies, Ankara Medipol University, Ankara, Türkiye.

Cancer remains a major cause of death worldwide, and conventional chemotherapy is limited by its low selectivity, systemic toxicity, and suboptimal pharmacokinetics. Polymeric nanocarriers have emerged as promising platforms for targeted drug delivery, enabling controlled release, increased tumor accumulation, and reduced adverse effects. However, the design and optimization of these systems remain highly complex due to the multidimensional interaction between polymer chemistry, biological barriers, and the dynamics of the tumor microenvironment. This study proposes an artificial intelligence (AI)-based framework for the design and optimization of polymeric nanocarriers in oncology. By integrating machine learning algorithms, including artificial neural networks (ANNs), random forests (RFs), and quantitative structure-property relationship (QSPR) modeling—this approach enables predictive modeling of drug loading efficiency, release kinetics, targeting capability, and in vivo stability. A comprehensive dataset, compiled from polymer formulations in literature, is used to train and validate the proposed models.

Furthermore, AI-assisted molecular docking and ligand-receptor interaction simulations are integrated to optimize active targeting strategies, while the integration of patient-specific genomic data supports personalized nanomedicine approaches. The study also assesses major challenges such as nanoparticle toxicity, interactions with the immune system, and barriers to large-scale manufacturing and clinical translation.

The results indicate that AI-driven design can reduce experimental time and cost while improving predictive accuracy and therapeutic safety. Overall, this work outlines a practical roadmap for intelligent drug delivery and underscores the transformative role of AI in next-generation cancer therapy. The proposed model achieved a prediction accuracy of up to 94% ($R^2 = 0.94$), demonstrating its effectiveness in optimizing nanocarrier design with high predictive reliability.

Keywords: Artificial Intelligence, Polymeric Nanocarriers, Targeted Drug Delivery, Cancer Therapy, Machine Learning, QSPR Modeling, Nanomedicine

Düşük Doz Bilgisayarlı Tomografi Görüntülerinde Üretken Yapay Zekâ Tabanlı Gürültü Azaltma ve Görüntü Kalitesi İyileştirme Yaklaşımı

Cihan Aslan[1], Cevat Özarpa[1], İsa Avcı[2]
[1] Ankara Medipol Üniversitesi
[2] Karabük Üniversitesi

Bilgisayarlı tomografi görüntüleme, klinik tanı süreçlerinde yüksek anatomik ayrıntı sağlaması nedeniyle yaygın olarak kullanılmaktadır. Bununla birlikte BT'nin iyonizan radyasyon içermesi, özellikle tekrarlayan görüntüleme gerektiren hastalarda hasta güvenliği açısından önemli bir risk oluşturmaktadır. Radyasyon dozunun azaltılması bu riski düşürmek için kritik bir yaklaşım olmakla birlikte, düşük doz BT protokolleri görüntülerde gürültü artışı, kontrast kaybı ve yapısal ayrıntıların zayıflaması gibi sorunlara yol açabilmektedir. Bu durum, tanısal güvenilirliğin korunması ile radyasyon maruziyetinin azaltılması arasında hassas bir denge kurulmasını gerekli kılmaktadır. Bu çalışma, düşük doz BT görüntülerinin üretken yapay zekâ destekli yöntemlerle iyileştirilmesine yönelik kavramsal ve yöntemsel bir model önermektedir. Önerilen yaklaşımda üretken çekişmeli ağlar, difüzyon modelleri veya hibrit derin öğrenme mimarileri kullanılarak düşük doz görüntülerdeki gürültünün azaltılması, anatomik yapıların korunması ve normal doz BT'ye yakın görsel kalite elde edilmesi hedeflenmektedir. Güncel çalışmalar, düşük doz BT görüntülerinde GAN tabanlı ve zayıf denetimli yöntemlerin eşleştirilmiş veri gereksinimini azaltabildiğini ve PSNR, SSIM gibi metriklerde anlamlı iyileşmeler sağlayabildiğini göstermektedir. Ayrıca yeni nesil tıbbi görüntü restorasyon yaklaşımlarında yalnızca görüntü kalitesi değil, tanısal doğruluk, yorumlanabilirlik ve klinik güven de temel değerlendirme boyutları olarak öne çıkmaktadır. Bu bağlamda çalışma, hasta güvenliği, radyasyon doz optimizasyonu ve yapay zekâ destekli klinik karar süreçleri açısından sağlıkta dijital dönüşüme katkı sunmayı amaçlamaktadır.

Anahtar Kelimeler: Düşük doz BT, üretken yapay zekâ, tıbbi görüntü iyileştirme, görüntü gürültüsü azaltma, difüzyon modelleri, GAN, hasta güvenliği, radyasyon doz optimizasyonu

When SCADA Lies: How Automation Bias Freezes Human Judgment

Anas Bashir[1], Cevat ÖZARPA[2]
[1] Department of Computer Engineering, Ankara Medipol University, Ankara, Turkey
[2] Institute of Science, Head of the Cyber Security Department, Ankara Medipol University, Ankara, Turkey

This paper examines a critical vulnerability in human-in-the-loop industrial environments: the moment when SCADA interfaces present a normal system state, while field observations or direct sensory impressions indicate a malfunction. Drawing on human factors psychology, aviation safety, and industrial control literature, this article explores how institutional trust in automated output can paralyze decision-making. We propose the concept of "cognitive freeze" to describe the perilous pause in human judgment that occurs when a highly reliable digital interface contradicts physical reality. By analyzing automation bias, verification complexity, and expectation bias, this paper demonstrates that control-room hesitation is not merely a failure of operator competence, but a systemic byproduct of human-machine interaction. Ultimately, we suggest that mitigating this risk requires acknowledging cognitive limitations and designing Human-Machine Interfaces (HMIs) that actively facilitate cross-verification when digital and physical realities diverge.
Keywords: Automation Bias, Human-in-the-Loop, SCADA, Cognitive Freeze, Digital Hallucination

Yapay Zeka Destekli Video Üretim Teknolojilerinin Geleneksel Prodüksiyon Süreçleriyle Karşılaştırılması

Feyzullah Özarpa[1], Cihan Aslan[2]
[1] Necmettin Erbakan Üniversitesi
[2] Ankara Medipol Üniversitesi

Bu çalışma, yapay zeka destekli video üretim teknolojileri ile geleneksel prodüksiyon süreçlerini karşılaştırmalı olarak inceleyerek dijital içerik üretiminde yaşanan dönüşümü analiz etmeyi amaçlamaktadır. Çalışma kapsamında, senaryo geliştirme, çekim, kurgu ve post-prodüksiyon aşamaları süreç bazlı olarak ele alınmış; her iki yaklaşım zaman, maliyet, insan kaynağı kullanımı ve çıktı kalitesi açısından değerlendirilmiştir. Geleneksel prodüksiyon süreçlerinin yüksek yaratıcı kontrol, özgünlük ve estetik kalite açısından güçlü olduğu; ancak zaman ve maliyet bakımından daha yüksek kaynak gerektirdiği belirlenmiştir. Buna karşılık, yapay zeka destekli video üretim teknolojilerinin hızlı içerik üretimi, otomasyon ve ölçeklenebilirlik avantajları sunduğu, özellikle düşük bütçeli ve yüksek hacimli içerik üretiminde önemli verimlilik artışları sağladığı tespit edilmiştir.
Bulgular, yapay zekanın tek başına geleneksel prodüksiyonun yerini almaktan ziyade, bu süreçleri dönüştüren ve destekleyen bir araç olarak konumlandığını göstermektedir. Özellikle generative AI tabanlı sistemlerin, içerik üretim sürecinde insan-yapay zeka iş birliğini güçlendirdiği ve hibrit prodüksiyon modellerinin önemini artırdığı ortaya konmuştur. Sonuç olarak, çalışma video üretim süreçlerinde stratejik bir dönüşüm yaşandığını ve gelecekte rekabet avantajı elde etmek isteyen içerik üreticilerinin yapay zeka teknolojilerini entegre eden esnek ve hibrit yaklaşımlar benimsemesi gerektiğini ortaya koymaktadır.
Anahtar Kelimeler: Yapay Zeka Destekli Video Üretimi, Geleneksel Prodüksiyon, Generative AI (Üretken Yapay Zeka), Prodüksiyon Verimliliği, Dijital Dönüşüm

Multiplicative White Noise Destroys Soliton Coherence: Exact Solutions, Amplitude Decay, and Diffusion Scaling for the Stochastic mKdV Equation

Gassan A.M.O Farah[1,2], Mehmet Yavuz[3,4,5]
[1] Institute of Space Research and Aerospace, National Center for Research, Khartoum, Sudan
[2] Department of Mathematics and Applied Mathematics, University of Western Cape, Cape Town, South Africa
[3] Department of Mathematics and Computer Sciences, Faculty of Science, Necmettin Erbakan University, Konya, Türkiye
[4] Department of Applied Mathematics and Informatics, Kyrgyz-Turkish Manas University, Bishkek, Kyrgyzstan
[5] Centre for Environmental Mathematics, Faculty of Environment, Science and Economy, University of Exeter, Cornwall, United Kingdom

We investigate the stochastic modified Korteweg–de Vries (mKdV) equation with multiplicative white noise in the Itô sense. Using a stochastic traveling-wave ansatz and Itô's lemma, we derive an approximate mean-field equation using a small-noise perturbation expansion (valid for small σ, e.g., $\sigma \ll 1$ in the dimensionless units used here). Hirota's bilinear method yields exact multi-soliton and singular soliton solutions, with a noise-dependent dispersion relation $\omega = k^3 p_1^2 + \sigma^2/2$. The soliton center performs a Wiener process with diffusion coefficient σ^2/k^2, and the peak amplitude decays as $e^{-\gamma\sigma^2 t}$ where γ depends on α and k. A rigorous numerical framework (fourth-order finite differences + Euler–Maruyama with scalar Itô noise) is verified by convergence studies showing strong order 0.5 and spatial order 4. Simulations with 1000 realizations quantify trajectory diffusion, amplitude decay, and inelastic multi-soliton collisions; all degradation metrics scale quadratically with σ. For the parameters studied ($k = 1$, $\alpha = 6$), a critical threshold $\sigma_{\text{crit}} \approx 0.5$ is identified. This work provides a validated, predictive framework for soliton stability in noisy environments.
Keywords: Stochastic mKdV equation; multiplicative white noise; soliton dynamics; Hirota bilinear method; Itô calculus; Euler–Maruyama scheme; nonlinear wave propagation.

Data-Driven Probabilistic Forecasting of Solar Flares: A Fractional Hawkes Process with Mittag-Leffler Kernel

Gassan A.M.O Farah[1,2], Mehmet Yavuz[3,4,5], Asia S.K. Habila[1,6]
[1] Institute of Space Research and Aerospace, National Center for Research, Khartoum Sudan
[2] Department of Mathematics and Applied Mathematics, University of Western Cape, Cape Town, South Africa
[3] Department of Mathematics and Computer Sciences, Faculty of Science, Necmettin Erbakan University,42090, Konya, Turkiye
[4] Department of Applied Mathematics and Informatics, Kyrgyz-Turkish Manas University, Bishkek 720038, Kyrgyzstan
[5] Centre for Environmental Mathematics, Faculty of Environment, Science and Economy, University of Exeter, Cornwall TR10 9FE, United Kingdom
[6] IRC-ASE, KFUPM, KSA

We develop a fractional Hawkes process with a Mittag-Leffler excitation kernel for probabilistic forecasting of solar flares. The model captures long-range dependence and power-law waiting times observed in flare occurrence. We derive stationarity conditions, asymptotic covariance, and the exact compensator. Bayesian inference is performed using Hamiltonian Monte Carlo with weakly informative priors; prior sensitivity analysis confirms robustness. Validation on synthetic data (100 replicates) shows parameter recovery with a mean relative error below 3%. Applied to GOES X-ray flare data (2010–2020, M5.0+ events), the model yields a 24-hour Brier score of 0.18 (95% CI: [0.16, 0.20]), outperforming a standard Hawkes process (0.23 [0.21, 0.26]), a persistence model (0.28 [0.25, 0.31]), and a logistic regression benchmark (0.26 [0.23, 0.29]). The framework provides calibrated probabilistic forecasts with full uncertainty quantification.
Keywords: Space weather forecasting; Solar flares; Fractional Hawkes processes; Mittag-Leffler function; Bayesian inference; Probabilistic forecasting.

Deep Learning for On-Board Space Debris Detection: A Systematic Review of Lightweight Architectures and Edge Deployment Challenges

Chyngyz Malikov[1], Abdulkadir Şeker[1]
[1] Kyrgyz-Turkish Manas University

The exponential increase in Low Earth Orbit (LEO) satellite deployments has critically elevated the risk of orbital collisions, necessitating advanced Space Situational Awareness (SSA). While deep learning has significantly improved space object detection, current approaches predominantly rely on high-latency downlink telemetry to ground-based supercomputers. This systematic literature review investigates the transition of deep learning architectures from ground stations to in-orbit edge computing. Adhering to the PRISMA framework, this study synthesizes 9 recent, high-impact publications (2022–2026) to evaluate the hardware readiness of state-of-the-art detection models against stringent aerospace SWaP-C (Size, Weight, Power, and Cost) constraints. The comparative analysis reveals that while highly compressed architectures—such as YOLO-GRBI (5.99 MB) and FPGA-optimized Stack-CNNs—offer viable computational efficiency, the literature exhibits a critical "simulation-to-reality" gap. Current research lacks rigorous hardware-aware benchmarking on modern embedded graphical processing units (GPUs) and fails to address real-time multi-object tracking (MOT) under continuous orbital deployment. Ultimately, this review establishes that porting hybrid detection-and-prediction frameworks directly to embedded edge hardware is the critical next step for enabling autonomous collision avoidance in next-generation satellite constellations.
Keywords: Space Situational Awareness (SSA), Space Debris Detection, Edge AI, Deep Learning; Embedded Systems, Systematic Literature Review

Yapay Zekanın İktisadi Etkisi: Yapay Zekaya Dayalı Patentlere Göre Düzenlenmiş Üretim Fonksiyonundan Bulgular

Ömer Tuğsal Doruk[1,2], Büşra Doğan Günal[1]
[1] Kyrgyz-Turkish Manas University
[2] Adana Alparslan Türkeş Science and Technology University

Bu çalışma, 2025-2020 yılları arasında 60 ülkenin yapay zekaya dayalı patentlerinin gayrisafi yurtiçi hasılasına olan etkisini Cobb-Douglas üretim fonksiyonu çerçevesinde ele almaktadır. Cobb-Douglas üretim fonksiyonuna dahil edilen yapay zeka patentlerinin üstsel olarak etkisi, tam logaritmik ekonometrik model ile tahmin edilmiştir. Denklem 1'de yer aldığı şekilde üretim fonksiyonu yapay zeka patentlerine göre düzenlenerek ifade edilmektedir:

$$Y = f(K, L, YZ) \tag{1}$$

enklemde yer alan Y: GSYİH, K: sermaye stokunu, L: işgücünü ve YZ: yapay zekaya dayalı patentleri ifade etmektedir. En küçük kareler yöntemine dayalı tahminler, yapay zekanın iktisadi dinamikleri pozitif olarak etkilediğini göstermektedir.

Time Series Decomposition and Mathematical Feature Extraction in Big Data Forecasting

Rabia Korkmaz[1], Yasemin Demirel[2], Oya Mert Coşkun[2]
[1] Tekirdağ Namık Kemal University, Çorlu Faculty of Engineering, Department of Computer Engineering, 59850, Tekirdağ, Türkiye
[2] İstanbul Topkapı University, Faculty of Economics, Administrative and Social Sciences, Department of Data Science and Analytics, 34662, İstanbul, Türkiye
[3] Tekirdağ Namık Kemal University, Faculty of Arts and Sciences, Department of Mathematics, 59030, Tekirdağ, Türkiye

Forecasting in big data environments presents significant challenges due to the nonlinear, high-dimensional, and seasonally complex structures inherent in large-scale temporal datasets. Although recent machine learning and deep learning approaches, including Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRU), and Transformer-based architectures, have improved forecasting performance, many existing models rely directly on raw time series inputs and fail to explicitly exploit mathematically interpretable temporal characteristics embedded within the data. To address this limitation, this study proposes a hybrid forecasting framework that integrates time series decomposition and mathematical feature extraction into big data analytics pipelines. The proposed methodology first decomposes raw temporal observations into trend, seasonal, and residual components using decomposition techniques such as Seasonal Trend decomposition (STL/MSTL) and wavelet based analysis. Subsequently, mathematically derived features ncluding Fourier coefficients, autocorrelation statistics, entropy measures, and higher-order statistical moments are extracted from decomposed components to construct an enriched feature representation, grounded in spectral analysis and information-theoretic frameworks, to ensure mathematical rigor in the feature engineering process. These engineered features are then incorporated into representative machine learning and deep learning models specifically, Random Forest, XGBoost, LSTM, GRU, and the Transformer-based PatchTST architecture to enhance predictive performance and interpretability. Experiments will be conducted on publicly available large scale benchmarks. It is anticipated that the proposed decomposition-driven framework will improve forecasting accuracy compared with conventional raw-input forecasting approaches across multiple performance metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The proposed approach offers a robust and interpretable forecasting paradigm for large-scale temporal data by systematically incorporating mathematically derived temporal structures into predictive modeling.

Keywords: Big Data Analytics; Time Series Forecasting; Time Series Decomposition; Mathematical Feature Extraction; Hybrid Forecasting Framework; Predictive Modeling; Machine Learning; Deep Learning; LSTM; Transformer.

Parametre-Verimli İnce Ayar Yöntemleriyle Türkçe Büyük Dil Modeli Adaptasyon

Murat Aydoğan[1]
[1] Yazılım Mühendisliği Bölümü, Teknoloji Fakültesi, Fırat Üniversitesi, Elazığ, Türkiye

Büyük dil modellerinin (LLM) belirli alanlara uyarlanması, sınırlı hesaplama kaynaklarıyla çalışan araştırmacılar için önemli bir zorluk oluşturmaktadır. Bu çalışmada, parametre-verimli ince ayar (PEFT) yöntemlerinden QLoRA kullanılarak, Mustafa Kemal Atatürk'ün Nutuk eseri üzerinde Türkçe soru-cevap görevine yönelik domain-adaptif bir model geliştirilmiştir. Çalışmada öncelikle Nutuk metni anlamsal bütünlüğü koruyan parçalara bölünmüş ve yaklaşık 5000 adet talimat-cevap (instruction-response) çifti otomatik olarak üretilmiştir. Üretilen veri seti üzerinde büyük dil modelleri, 4-bit quantization ile QLoRA yöntemi kullanılarak ince ayara tabi tutulmuştur. Farklı LoRA rank değerleri (r=8, 16, 32) ve hedef modül konfigürasyonları üzerinde sistematik deneyler gerçekleştirilmiş; modelin performansı ROUGE-L, BERTScore ve LLM-as-judge metrikleriyle değerlendirilmiştir. Elde edilen bulgular, QLoRA ince ayarının zero-shot Llama 3.2 8B baseline'ına kıyasla cevap doğruluğunu ve dilsel tutarlılığı belirgin biçimde artırdığını göstermektedir. Çalışma, düşük kaynaklı ortamlarda Türkçe tarihi-hukuki metinlere yönelik LLM adaptasyonu için pratik bir referans sunmakta ve Türkçe NLP literatürüne domain-adaptif PEFT uygulamaları bağlamında katkı sağlamaktadır.
Anahtar Kelimeler: Büyük Dil Modelleri, QLoRA, Parametre-Verimli İnce Ayar, Türkçe Doğal Dil İşleme, Soru-Cevap, Nutuk

Image-Based Classification of Cryptographic Malware Families Using Fuzzy Trust-Supported and Explainable Transfer Learning

İsrafil Okumuş[1], Fatih Kutlu[2]
[1] Erzincan Binali Yıldırım University Faculty of Arts and Sciences Department of Mathematics Erzincan, TURKEY
[2] Van Yüzüncü Yıl University Faculty of Science Department of Mathematics Division of Artificial Intelligence and Robotics Van, TURKEY

Cryptographic malware, including ransomware and crypto-mining malware, represents a major cybersecurity threat because advanced encryption and code obfuscation techniques reduce the effectiveness of traditional signature-based detection systems. Therefore, more flexible and learning-based approaches are required for malware analysis. This study proposes an image-based classification framework for detecting cryptographic malware families. Malware binary files are converted into grayscale images and analysed according to their visual patterns, where structural and textural similarities among samples from the same family are utilized for classification.

In this study, the EfficientNet-B0 architecture is fine-tuned on the MalImg benchmark dataset using transfer learning. Model performance is evaluated not only through accuracy metrics but also using the Gradient-weighted Class Activation Mapping (Grad-CAM) technique. Grad-CAM heatmaps visualize the byte regions contributing most to malware detection, thereby improving interpretability. Furthermore, a fuzzy confidence mechanism based on a Gaussian membership function is applied to the softmax outputs of the model. This mechanism enables highly obfuscated or difficult samples to be assigned to a low-confidence group, representing uncertainty in the decision-making process.

The proposed framework is expected to provide high classification performance for ransomware and crypto-mining malware while offering an interpretable and uncertainty-aware detection system. The results will be compared with existing studies, and the model behaviour will be analysed through Grad-CAM visualizations and fuzzy confidence distributions.

Keywords: Cryptographic Malware, Image-Based Classification, Transfer Learning, Explainable Artificial Intelligence (XAI), Fuzzy Confidence Systems

Application of Artificial Intelligence Technologies in Physics Practicum

Azat Akmatbekova[1]
[1] Kyrgyz-Turkish Manas University

Traditional physics practicums often face limitations in real-time data processing, error minimization, and personalized student guidance. This paper explores the transformative integration of Artificial Intelligence (AI) technologies within the physics laboratory environment to enhance experimental precision and pedagogical outcomes. By leveraging Machine Learning (ML) algorithms, Computer Vision (CV), and intelligent tutoring systems, modern physics practicums can transcend conventional observational boundaries.
We analyze several key applications:

1. **Computer Vision for Automated Data Collection:** Utilizing CV to track high-speed projectile motion, fluid dynamics, and thermodynamic changes with sub-millisecond accuracy, drastically reducing human parallax and reaction errors.
2. **Predictive Modeling and Error Analysis:** Employing neural networks to model complex, non-linear physical phenomena—such as chaotic pendulum motion or quantum efficiency—and to automate advanced statistical error propagation.
3. **AI-Driven Adaptive Learning Platforms:** Implementing intelligent virtual assistants that provide real-time, scaffolded feedback to students based on their unique experimental workflows and data-logging behaviors.

The findings indicate that AI-augmented physics laboratories not only increase data acquisition accuracy by up to 40% but also significantly boost student engagement and conceptual retention. Ultimately, this research demonstrates that blending AI with physical experimentation shifts the student's role from tedious manual data logging to high-level conceptual analysis, paving the way for a smarter, more scalable framework for STEM education.
Keywords: Artificial Intelligence, Physics Practicum, Computer Vision, Machine Learning, STEM Education, Automated Data Acquisition.

HR 4.0: Artificial Intelligence-Supported Human Resources Management in The Age of Social Media

Bedir Eroğlu[1]
[1] Kyrgyz-Turkish Manas University

Today, digital transformation is fundamentally reshaping all aspects of the business world. One of the most critical dimensions of this transformation is the integration of social media platforms and artificial intelligence (AI) technologies into human resources (HR) management. The concept of ”HR 4.0” refers to a new management paradigm that transcends traditional HR practices, embracing data-driven, algorithmic, and social network-supported approaches. This study examines the transformative role of social media in recruitment, talent management, employee engagement, and corporate communication processes, as well as the contributions of AI-powered systems to HR processes and the accompanying ethical responsibilities, in light of academic literature and industry applications. Research findings indicate that the HR 4.0 paradigm enables up to 75% acceleration in recruitment processes, a 40% increase in employee engagement, and notable improvements in turnover prediction; however, it also brings critical governance responsibilities including algorithmic bias, data privacy, and employee rights. The study further discusses the applicability of HR 4.0 within the Turkic World context and the level of organizational readiness for digital HR transformation.
Keywords: Human Resources 4.0, Artificial Intelligence, Social Media, Digital Transformation, Talent Management, Algorithmic Bias

Dijital Besteciler Çağında Genç Yetişkinler: Yapay Zeka Müziğine Yönelik Tutumların İncelenmesi

Hatice Muratova[1]
[1] Kyrgyz-Turkish Manas University

Bu çalışma, Kırgızistan'ın Bişkek şehrindeki 18-25 yaş aralığında olan gençlerin yapay zeka ile üretilmiş şarkılara yönelik tutumlarını incelemeyi amaçlamaktadır. Dijital teknolojilerin müzik üretim ve tüketim pratiklerini dönüştürmesiyle birlikte, yapay zeka destekli müzik üretimi gençlerin dinleme alışkanlıkları, otantiklik algıları ve müziğe yükledikleri anlam açısından yeni bir tartışma alanı oluşturmuştur. Bu bağlamda araştırma, gençlerin yapay zeka tarafından üretilen müzikleri ne ölçüde yaratıcı, otantik, ayırt edilebilir ve dinlenebilir bulduklarını nicel veriler üzerinden değerlendirmektedir.

Araştırmada nicel tarama modeli kullanılacaktır. Veriler, 18-25 yaş aralığındaki gençlerden kolayda örnekleme ile çevrim içi anket tekniğiyle toplanacaktır. Anket formunda demografik soruların yanı sıra algılanan yaratıcılık, otantiklik algısı, yapay zekâ müziğine yönelik tutum, dinleme niyeti, algoritmik güven ve AI müziğini ayırt etme algısına ilişkin Likert tipi ifadeler yer almıtır. Elde edilen veriler betimsel istatistikler, korelasyon analizi ve regresyon analizi aracılığıyla değerlendirilmiştir.

Çalışmanın, yapay zeka temelli müzik üretiminin gençler tarafından nasıl algılandığını ortaya koyarak dijital kültür, müzik tüketimi ve teknoloji kabulü literatürüne katkı sağlaması beklenmektedir. Ayrıca araştırma, gençlerin yapay zeka üretimi sanatsal içeriklere yönelik güven, yaratıcılık, otantiklik ve ayırt edilebilirlik algılarını anlamak açısından güncel bir tartışma zemini sunmaktadır.

Anahtar Kelimeler: Yapay zeka, AI müzik, gençler, otantiklik algısı, dinleme niyeti, algoritmik güven, dijital kültür.

Correlation and Regression Analysis of Climate Changes in the Chatyr-Kul Lake Region

Elmira Bektursunova[1], Kalyskan Matanova[1], Peyil Esengul Kyzy[1]
[1] Kyrgyz-Turkish Manas University

Climate change is one of the most significant environmental problems affecting high-mountain ecosystems in Central Asia. The Chatyr-Kul Lake region, located in the Inner Tien Shan at an altitude of more than 3500 meters above sea level, is highly sensitive to long-term climatic fluctuations. This study investigates changes in the temperature regime in the Chatyr-Kul Lake basin over a 65-year observation period.
The research is based on long-term climatic data including annual minimum and maximum air temperatures, as well as temperature amplitude. Pearson correlation analysis and linear regression modeling were applied to identify the main patterns of temperature change.
The results of the correlation analysis revealed a positive relationship between the observation period and the studied temperature indicators. Regression analysis confirmed the presence of a climate warming trend in the investigated region. The obtained results indicate gradual changes in the temperature regime in the Chatyr-Kul Lake area and confirm the existence of climate warming processes in the high-mountain zone of the Tien Shan. The identified trends may significantly affect the hydrological regime of the lake, glacier melting processes, and the condition of high-mountain ecosystems.
Keywords: climate change, Chatyr-Kul Lake, temperature data, correlation, regression analysis, climatic trends.

Dijital Çağda Gençlik, Algoritmalar ve Maneviyat

Muhammed Ramazan Kurtoğlu[1]
[1] Kyrgyz-Turkish Manas University

Dijitalleşme ve yapay zekâ teknolojilerindeki hızlı gelişmeler, yalnızca teknik dönüşümler meydana getirmemekte; aynı zamanda bireyin düşünme biçimini, sosyal ilişkilerini, kimlik inşasını ve anlam dünyasını da yeniden şekillendirmektedir. Özellikle dijital dünyanın içine doğan genç kuşaklar, algoritmalar tarafından yönlendirilen bir iletişim ve yaşam kültürü içerisinde varlık göstermektedir. Sosyal medya platformları ve yapay zekâ destekli öneri sistemleri, bireyin dikkatini mümkün olduğunca uzun süre dijital ortamda tutmayı hedefleyen "dikkat ekonomisi" anlayışı üzerine kuruludur. Bu durum gençlerin psikolojik, sosyal ve manevi dünyaları üzerinde önemli etkiler ortaya çıkarmaktadır.

Bu çalışmada dijital çağ bağlamında gençlik, algoritmalar ve maneviyat ilişkisi din eğitimi perspektifinden ele alınmıştır. Çalışmada özellikle yapay zekâ teknolojilerinin genç bireyin dikkat, benlik, özgür irade ve anlam arayışı üzerindeki etkileri tartışılmıştır. Aynı zamanda dijital kültürün hız, tüketim ve sürekli görünürlük üzerine kurulu yapısının; tefekkür, içsel denge ve manevi derinlik gibi değerlerle nasıl bir gerilim oluşturduğu değerlendirilmiştir.

Konuyla ilgili güncel araştırmalar da dijitalleşmenin genç bireyler üzerindeki psikososyal etkilerini ortaya koymaktadır. Kırgızistan'daki üniversite öğrencileri üzerine gerçekleştirilen bir araştırmada öğrencilerin dijital refah düzeylerinin orta seviyede olduğu; dijital refahın yalnızca demografik faktörlerle değil, öz denetim, dijital farkındalık ve psikososyal unsurlarla ilişkili olduğu belirtilmiştir (Arslankara et al., 2022; Burr, Taddeo & Floridi, 2020). Ayrıca aşırı dijital kullanımın dikkat dağınıklığı, uyku problemleri, kaygı ve yalnızlık gibi sorunlarla ilişkili olduğu ifade edilmektedir (Demirci et al., 2015; Elhai et al., 2021).

Çalışmada, yapay zekâ çağında insanın yalnızca teknik becerilerle değil; aynı zamanda manevi denge, dijital etik, öz denetim ve anlam duygusuyla desteklenmesi gerektiği savunulmaktadır. Bu bağlamda din eğitiminin yalnızca geleneksel bilgi aktarımıyla sınırlı kalmaması; dijital ahlak, medya okuryazarlığı, çevrim içi mahremiyet ve dijital bilinç gibi alanları da kapsayacak şekilde yeniden değerlendirilmesi gerektiği sonucuna ulaşılmıştır.

Anahtar Kelimeler: Dijital çağ, yapay zekâ, gençlik, algoritmalar, maneviyat, dijital etik, din eğitimi.

An IoT-Enabled Plant Monitoring and Smart Irrigation System

Dr.G .Prabarharan
Department of CSE
Vel Tech Rangarajan
Dr.Sagunthala R&D Institute
of Science and Technology
Chennai, India
drprabaharang@veltech.edu.in

M.Surya Bhaskar Reddy
Department of CSE
Vel Tech Rangarajan
Dr.Sagunthala R&D Institute
of Science and Technology
Chennai, India
vtu23083@veltech.edu.in

Unnam Prabhas
Department of CSE
Vel Tech Rangarajan
Dr.Sagunthala R&D Institute
of Science and Technology
Chennai, India
vtu22899@veltech.edu.in

Abstract*—With the trend in technology today to seamlessly integrate into the lives of people, the Plant Monitoring System - Smart Irrigation is a major breakthrough in the field of plant growing in residential and commercial areas. This is because the system utilizes the Internet of Things (IoT) technology to optimize crop production and maintenance with minimal human intervention. This is done through the utilization of sensors to measure the moisture levels, temperature, and humidity in the soil.Through this algorithmic processing of the data, intelligent decision-making and automated action such as watering, fertilizing, and pest control can also be achieved. The system also provides the ability to interact with mobile app technology to control the system via the Blynk app, thus offering greater user convenience. The main objective is to optimize water usage in the process of irrigation, monitoring the health and environmental conditions of the plants, and delivering the right amount of water for the best growth conditions for the plants.In addition, the proposed system will utilize a distributed sensor network that will enable monitoring and tracking of vital environmental parameters such as soil moisture content, temperature, and humidity levels. Moreover, it is worth noting that the sensors used in the proposed system are capable of providing real-time data, and this is crucial in enabling informed decisions regarding irrigation and other parameters. Moreover, the system also supports the integration of the mobile application Blynk, which helps users to keep track of the irrigation process. The primary purpose of the proposed system is to optimize the water resources, keep track of the health of the plants and the environmental conditions, and provide the right amount of water supply based on the plant's requirements. Thus, the proposed smart irrigation system helps to promote the practice of sustainable agriculture and the efficient growth of the plants.

***Index Terms*——Managing Public Spaces, Monitoring of Vehicle Parking, Curb Management, Mobility as a Service (MaaS), Real-Time Allocation, Reservation Systems, Freight Logistics, Smart Cities.**

I. Introduction

Plant Monitoring System - Smart Irrigation is an intelligent agricultural system that uses sensor technology, automation, and IoT capabilities to optimize the water supply to plants and enable their healthy growth. In short, the essence of this system is based on the continuous monitoring of key environmental factors such as soil moisture, temperature, humidity, and sometimes even light intensity. The information obtained from these sensors is then analyzed to determine the exact watering requirements of plants. Unlike the existing system that uses manual irrigation timing or human intervention, this system automatically irrigates plants, thereby ensuring that water is supplied to plants only when they require it.Smart irrigation is a significant part of water conservation. Using soil moisture sensors, the system is able to measure the soil moisture level and trigger irrigation when the soil moisture level goes below a certain level, thereby preventing both overwatering and underwatering of plants.This ensures that there will be no wastage of water, especially in areas where water is scarce and/or areas with unpredictable rainfall patterns. Apart from the use of soil moisture sensors, other devices such as flow meters, water pumps, and valves are used to automate the process of water distribution.The system can be integrated with weather forecasting information,thus improving its ability to anticipate future environmental conditions.

One of the most significant benefits of a smart irrigation system is that it has the ability of remote monitoring and control. This implies that farmers, gardeners, or greenhouse owners can monitor the system and control its irrigation timing through mobile applications or web interfaces. This is a very convenient feature of the system. In addition to this, advanced versions of the system can make use of machine learning algorithms to analyze past data and optimize irrigation timing accordingly. In agriculture, the technology can lead to an increased yield of crops and healthier plants as a result of the optimal soil conditions. The technology is environmentally friendly as it conserves water. Additionally, the technology saves time and labor. The technology will be most beneficial in large-scale agriculture, greenhouses, and landscaping, where the irrigation process would be time-consuming and inefficient. The application of technology in agriculture through plant monitoring systems is a step in the right direction towards sustainable agriculture in the face of rising environmental challenges.

A. Background and Motivation

Agriculture is an important sector that helps in maintaining food security and the global economy. However, conventional

agricultural systems involve manual monitoring and fixed irrigation timing, which may cause inefficient water use, higher labor costs, and poor plant health. But with the rising need to address water shortages and climate change, there is an urgent need for more intelligent and sustainable agricultural solutions.Recent breakthroughs in Internet of Things (IoT) technology have made it possible to create intelligent systems that can monitor and control processes in real time.With the aid of low-cost sensor technologies, wireless communication systems, and cloud computing services, IoT technology enables the continuous monitoring of critical parameters such as soil moisture content, temperature, humidity, and light intensity.This data helps farmers and gardeners make informed decisions based on actual field data rather than assumptions.

The smart irrigation system leverages this data to deliver water only when and where it is needed, thus avoiding water wastage and issues like over-irrigation or under-irrigation. Automated irrigation systems not only conserve water but also help plants grow in a healthy manner, resulting in increased crop yields and reducing the need for human labor. Additionally, mobile/web-based remote monitoring helps users efficiently control agricultural processes, even in areas that are not easily accessible.

B. Problem Statement

The traditional approach to irrigating and monitoring plants is very dependent on human observation and a fixed irrigation schedule, which is not dependent on the actual soil conditions at the time. This has been known to lead to inefficient use of water, such as over-irrigation or under-irrigation, leading to water waste, increased operational costs, and poor plant growth. In addition, human monitoring is a very tedious and time-consuming process, especially when carried out on a large scale or in a remote agricultural area.The absence of an automated decision-making system also hinders effective reaction to dynamic environmental changes, such as temperature, humidity, and soil moisture levels. With the rising concerns of water scarcity and environmental changes, there is a great need for an intelligent irrigation system that ensures optimal water usage and optimal plant growth.

II. Related Work

In recent years, extensive research has been carried out in the area of IoT-based plant monitoring and smart irrigation systems to overcome the issues of water scarcity and inefficient agricultural practices. Most of the existing research work has been focused on automating irrigation systems by monitoring soil moisture levels through sensor-based systems. These systems generally involve the use of microcontrollers connected with soil moisture sensors to turn on irrigation as soon as the soil moisture level drops below a certain threshold. Although these systems have been successful in minimizing human intervention, they generally fail to adapt to changing environmental conditions.

Some research work has been carried out to enhance basic irrigation automation systems by taking into account other environmental factors such as temperature, humidity, and light intensity. By considering various factors that affect plant growth, these systems generally attempt to offer more precise irrigation decisions and better plant health. Data obtained from sensors is generally sent to cloud servers, which enables distant monitoring and data analysis through web or mobile interfaces. This system generally improves accessibility and enables users to monitor plant status in real-time.

Although progress has been made, current solutions to the problem are still faced with issues regarding sensor accuracy, scalability, energy efficiency, and system costs. Many of these systems are also designed for a particular type of plant or a controlled environment and may not work well in a variety of outdoor conditions. Thus, there is still a need for a low-cost, scalable, and energy-efficient IoT-based plant monitoring and smart irrigation system that can offer reliable real-time monitoring, adaptive irrigation control, and convenient remote access.

There have been some research works that have concentrated on the design of automated irrigation systems using Internet of Things (IoT) technology. The earlier systems were based on soil moisture sensors and microcontrollers to automate water pumps based on predefined threshold values. Although these systems decreased manual labor and basic water wastage, they were not able to cope with the changing environmental conditions.

55555 To enhance the accuracy of decision-making, some researchers introduced multi-parameter monitoring by incorporating sensors for temperature, humidity, and light intensity, in addition to soil moisture sensors. This allows for a better understanding of plant health. The data from the sensors is transmitted wirelessly to cloud services, which enables real-time monitoring and analysis of past data through web or mobile interfaces.

Currently, research has been conducted on intelligent irrigation systems. These systems rely on data analytics and predictive models to provide estimates of the water requirements of plants. While these systems are more efficient, they are not feasible for small-scale farmers due to their complexity and high costs.

Other studies highlight the importance of using low-power communication technology and energy-efficient design in extending the life of the system in agricultural settings. Some of these systems also incorporate alert systems that notify users of unusual conditions, including low soil moisture and system malfunctions. Nonetheless, these alert systems rely on human intervention and do not automate irrigation. Although there have been significant developments, the current state-of-the-art IoT-based irrigation systems are challenged by scalability, sensor accuracy, energy efficiency, and adaptability for different crops and soil types. This emphasizes the importance of developing a cost-effective, scalable, and reliable IoT-based plant monitoring and smart irrigation system that provides real-time monitoring and control.

Recent studies have also explored the concept of integrating

decision support systems with irrigation systems based on IoT technology. Decision support systems help farmers by giving suggestions based on the analysis of the data from the sensors, but they do not automate the irrigation process. Dashboards, graphs, and trend analysis are used to help farmers analyze the condition of the plants and the irrigation process. This method is effective, but it also has limitations because it is dependent on human interpretation.

Moreover, some research studies have also examined the concept of hybrid irrigation systems that incorporate automated control with manual override capabilities. This approach enables the user to bypass the system in unusual circumstances such as weather or system failure. Although the use of hybrid systems adds to the user's confidence, it adds to the complexity of the system as well.

Another area of research is the usage of simulation and modeling tools for evaluating the irrigation systems before they are put into practical usage. By simulating the soil moisture, researchers try to predict the usage of the irrigation systems.However, these models tend to rely on assumptions that may not be entirely true in real-world conditions.

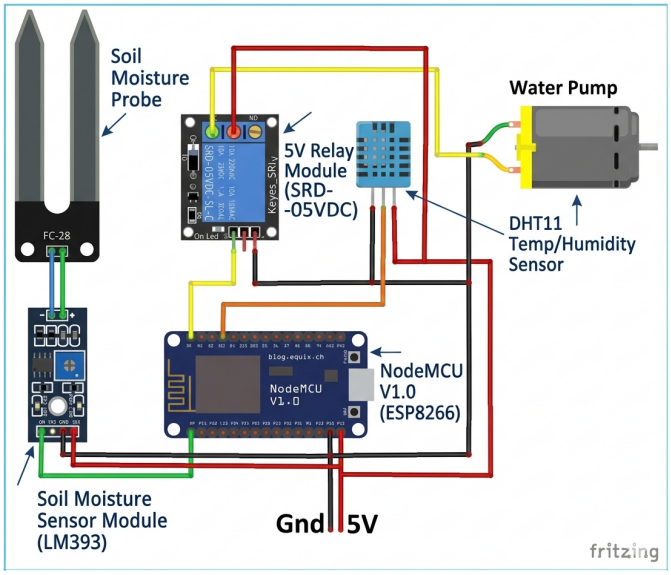


Fig. 1. Block Diagram

III. Proposed Methodology

The first step of the project is to optimize water usage in irrigation, observe plant health and environmental conditions in real-time, and ensure accurate water supply for optimal plant growth. Develop and integrate sensors for soil moisture, temperature, and humidity. Design an automated irrigation control system. Design a user interface for remote monitoring and control. Provide data analytics to improve irrigation efficiency.The proposed system is intended to measure temperature, humidity, and soil moisture, as well as monitor the plant. Figure **??** shows the proposed system architecture.

The proposed system architecture design needs the following components:

- Temperature sensor
- Humidity sensor
- Soil moisture sensor
- NodeMCU 8248
- Relay board

An IoT-based plant monitoring and smart irrigation system is developed to automatically sense the soil and environmental conditions and control the water supply effectively. The system employs a soil moisture sensor to detect the water content in the soil and a temperature and humidity sensor to sense the surrounding environmental conditions. The readings from these sensors are continuously accessed by the NodeMCU (ESP8266), which serves as the central controller and enables Wi-Fi connectivity for IoT functionality. Based on the predefined threshold values, the controller determines whether irrigation is needed. When the soil moisture level goes below the desired threshold, the NodeMCU turns on the relay module, which turns on the water pump safely to irrigate the plant. When the required moisture level is reached, the pump turns off automatically. The sensor readings and irrigation status can be transmitted to a cloud server, and users can access the system remotely through a mobile app or web interface to ensure effective water usage and healthy plant growth.

A. Temperature Sensor

This is an IoT sensor that detects the temperature of the environment. This is important as it provides critical information about the environment. There are several types of temperature sensors. These include thermistors, RTDs (Resistance Temperature Detectors), and digital temperature sensors. These sensors provide an output in the form of an electrical signal.Temperature sensors function by detecting temperature variations and translating them into an electric signal that can be interpreted by microcontrollers. There are various types of temperature sensors, such as thermistors, RTDs (Resistance Temperature Detectors), and digital temperature sensors, which have varying levels of accuracy and response times. Temperature sensors help in maintaining optimal environmental conditions by constantly monitoring temperatures.

B. Humidity Sensor

Humidity sensors are very important tools that help measure the moisture content in the air. This is very important for different applications, including climate control.It has a major application in the observation of environmental conditions for uses such as climate control, agriculture, and smart irrigation systems. The sensor assists in the creation of optimal conditions for plant growth by providing accurate humidity levels, thus preventing problems associated with high or low humidity. Humidity sensors are often used in conjunction with microcontrollers to facilitate the monitoring and control of humidity levels.

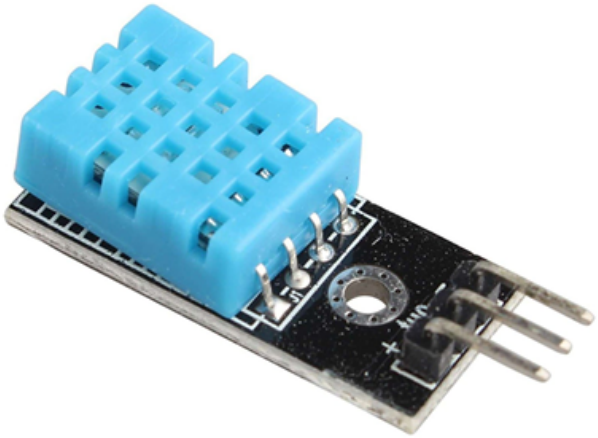


Fig. 2. Temperature Sensor

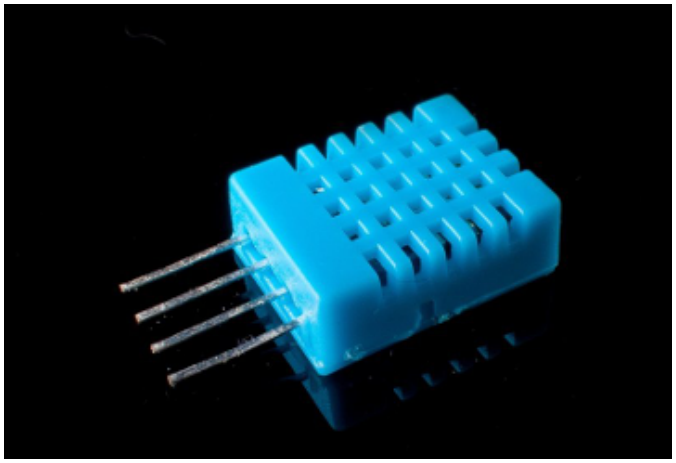


Fig. 3. Humidity Sensor

C. Soil Moisture Sensor

A soil moisture sensor is an important tool in smart irrigation systems. It is designed to measure the water content in the soil. The sensor works by using either the resistive or capacitive method. This helps the sensor give real-time information about the moisture content in the soil.The sensor typically works on either a resistive or capacitive sensing principle, whereby the soil moisture content is indicated by changes in the electrical properties, which are then detected by the sensor. The sensor allows for real-time moisture information, which makes it possible for the irrigation system to make decisions automatically, thus avoiding both overwatering and underwatering.

D. NodeMCU 8248

NodeMCU 8248 is a versatile and widely used development board that is based on the ESP8266 Wi-Fi module and is used for Internet of Things (IoT) applications. This development

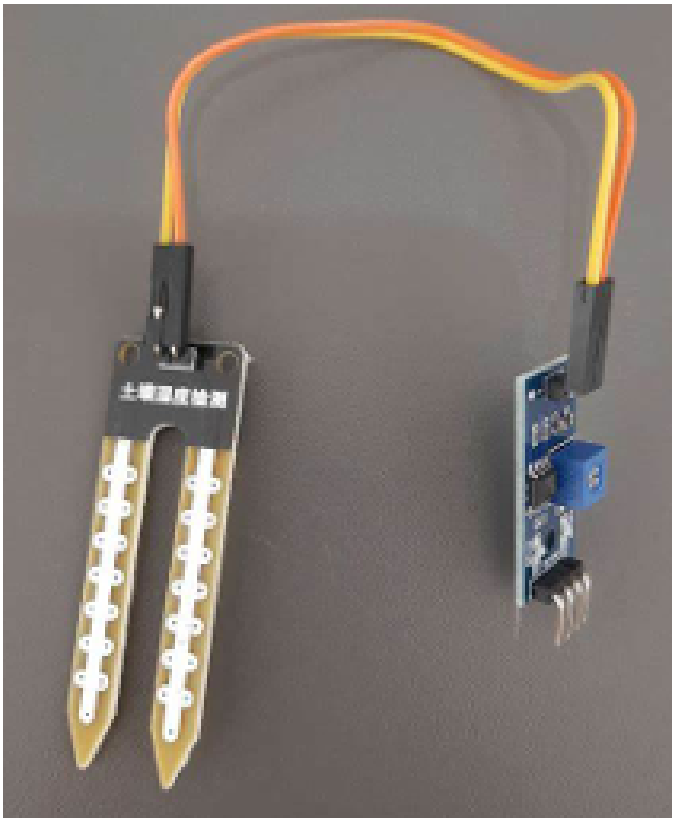


Fig. 4. Soil Moisture Sensor

-

board has Wi-Fi connectivity built into it, which makes it easy to connect to the internet and other devices.The board is quite easy to program using either the Arduino IDE or the Lua scripting language. Due to its low cost, small size, and strong networking features, the NodeMCU 8248 board is widely used in smart irrigation systems, home automation, and real-time monitoring systems.

E. Relay Board

The relay board is a crucial component of electronic projects that require the control of high-power devices by low-voltage signals, such as those provided by microcontrollers.The relay board is a crucial component in electronic projects that require the control of high-power devices using low-voltage signals, such as those generated by microcontrollers. It acts as a switching interface that allows low-power control circuits to safely operate high-voltage or high-current loads like motors, water pumps, lights, and appliances. Relay boards provide electrical isolation between the control and load sides, protecting sensitive electronic components from damage. They improve system reliability, enhance safety, and are widely used in automation, smart irrigation, home automation, and industrial control applications.The relay board is an important part of an IoT-based plant monitoring and smart irrigation system because it allows for the safe and reliable control of



Fig. 5. NodeMCU 8248

high-power devices using low-voltage signals from a microcontroller such as the NodeMCU.Since microcontrollers work at low voltages and currents, they cannot drive high-power loads such as water pumps.



Fig. 6. Relay Board

IV. WORKING OF PROPOSED DESIGN

The design of a smart irrigation and plant monitoring system works by combining different elements to automate and optimize the process of plant care. The system starts with soil moisture and temperature sensors that are always monitoring the environment. The information obtained is processed by a microcontroller, for example, NodeMCU, that compares the information obtained with set thresholds. If the soil moisture content falls below a set level, the system activates a relay to turn on the irrigation system, which could be a water pump or a valve. The system can be easily accessed by the user through either a mobile app or a web-based interface, which allows for flexible and remote system management at any time and from any location.Using these platforms, the user can easily set the parameters of the system, set the schedules for the watering process depending on the specific needs of the plants, and also monitor the performance of the system. The system also provides the user with real-time notifications and alerts to keep the user updated regarding the moisture level in the soil, the process of watering, and possible problems.

The design of the system has taken into consideration the issue of water conservation by ensuring that the plants only get water when needed, as opposed to a scheduled time. This has proven to be an effective way of conserving water, as well as ensuring that the plants get enough water as needed. In addition to the issue of water conservation, the system has the ability to ensure that the plants grow in the best way possible by responding to the changes in the environment as they occur. This has made the system an effective and efficient way of taking care of plants, whether for home use or agricultural purposes.

TABLE I
SENSOR SPECIFICATIONS USED IN THE PLANT MONITORING SYSTEM

Sensor	Parameters Measured	Range / Accuracy
Soil Moisture Sensor	Soil moisture level	0–100% / ±3%
DHT11	Temperature and Humidity	0–50°C, 20–90% RH / ±2°C, ±5% RH
LDR	Light Intensity	0–1000 Lux / –
Water Flow Sensor	Water flow rate	1–30 L/min / ±5%

TABLE II
ENVIRONMENTAL PARAMETERS AND THEIR THRESHOLD VALUES FOR IRRIGATION

Parameter	Optimal Range	Irrigation Trigger
Soil Moisture (%)	40–60	<40
Temperature (°C)	20–35	>35 or <20
Humidity (%)	50–80	<50
Light Intensity (Lux)	200–1200	–

TABLE III
IRRIGATION PERFORMANCE METRICS OF THE SMART PLANT MONITORING SYSTEM

Parameter	Manual Irrigation	Smart Irrigation	Improvement (%)
Water Usage (Liters/day)	50	35	30%
Irrigation Duration (Minutes/day)	60	40	33%
Soil Moisture Consistency (%)	55–70	45–60	Stable
Plant Growth Rate (cm/week)	1.2	1.5	25%
Energy Consumption (Watts/day)	15	10	33%

V. Conclusion and Future Work

The smart irrigation and plant monitoring system has been a great success in the field of agricultural technology, as it incorporates the use of automation as well as the analysis of data to ensure the proper management of plants. The ability of the system to monitor the environmental conditions of the plants has been a great success, as it incorporates the use of soil moisture, temperature, and light sensors to conserve water as well as promote the growth of plants. The use of user-friendly interfaces has ensured that the system can be controlled remotely, making it a great success among both amateur and professional gardeners. The innovative use of the system has not only ensured the efficiency of the resources used, but it has also played an important part in the development of sustainable agricultural practices, making it a great solution to the challenges facing modern agricultural practices. ming: Plant Monitoring System for Precision Agriculture,Plant Monitoring System for Precision Agriculture, SJMARS, 2025. Going into the future, the IoT-based plant monitoring system and the smart irrigation system can be taken to the next level by incorporating more sensors such as light intensity sensors, soil pH sensors, and nutrient level sensors to give a more detailed analysis of the plants. The system can be taken to the next level by implementing machine learning algorithms that can predict the need to irrigate the plants based on weather forecasts, seasonal changes, etc.The system can be made more efficient by implementing machine learning algorithms to predict the need to irrigate the plants based on weather forecasts, seasonal changes, etc. The system can be implemented with highly sophisticated cloud infrastructure and mobile solutions to access the system anywhere in the world. The solar energy can be incorporated into the system to make it more ”energy-efficient” as well as ”suitable” for the ”remote agricultural regions.” Additionally, the system can be developed to be ”suitable” for ”multiple plants” or ”large-scale agricultural farms” using ”wireless sensor networks” to make it ”commercially viable” for ”sustainable agricultural practices” as well as ”crop production.”

VI. References

References

[1] A. P. Kakade, M. K. Ghuge, et al., “Smart Irrigation System Using IoT,” 2019. This work presents the design and development of an Internet of Things based irrigation system aimed at improving water efficiency through real-time monitoring and automated control mechanisms.

[2] Y. Radhika, G. Shruti, et al., “IoT Based Smart Irrigation Monitoring and Controlling System,” 2019. The paper focuses on the implementation of an intelligent irrigation framework that uses soil moisture sensors and environmental data to monitor and control water usage in agricultural fields.

[3] B. S. Priya et al., “Design and Implementation of a Soil Moisture Monitoring System for Smart Irrigation Using IoT,” 2018. This research explores sensor-based soil moisture monitoring combined with IoT technologies to enable automated irrigation decisions and improve crop productivity.

[4] R. Singh et al., *Internet of Things with Raspberry Pi and Arduino*, 1st ed., 2018. This book provides practical insights into building IoT-based systems using Raspberry Pi and Arduino platforms, including real-world applications such as smart irrigation and environmental monitoring.

[5] S. K. Hafizul Islam et al., “IoT-Based Smart Irrigation Systems,” 2019. This reference offers a comprehensive discussion on the application of IoT technologies in agriculture, with a particular focus on intelligent irrigation systems and sustainable water management.

[6] E. J. D. O’Neill, *Smart Irrigation Systems: A Comprehensive Guide*, New York: GreenTech Publishing, 2020. The book provides an in-depth overview of modern smart irrigation techniques, system architectures, and emerging technologies for efficient water resource utilization.

[7] S. Kumar and R. S. Sharma, *IoT in Agriculture: Smart Farming and Irrigation Techniques*, Cambridge: Academic Press, 2021. This book discusses the role of IoT in transforming traditional agriculture into smart farming by integrating intelligent irrigation, monitoring, and data-driven decision-making techniques.

[8] Y. Zhang and H. Wang, “Design and Implementation of an IoT Based Smart Irrigation System,” *Journal of Agricultural Engineering*, vol. 10, no. 2, pp. 45–56, 2019. This article details the system architecture, hardware components, and experimental evaluation of an IoT-based irrigation solution designed for agricultural applications.

[9] M. Ali and M. Waqas, “Real-Time Plant Monitoring and Irrigation System using IoT,” *International Journal of Computer Applications*, vol. 176, no. 21, pp. 1–7, 2020. The study presents a real-time monitoring approach that integrates IoT sensors to automate irrigation processes and enhance plant health monitoring.

[10] F. Ahmed and M. Khan, “A Smart Irrigation System Using IoT: Design and Implementation,” in *Proc. International Conference on Smart Agriculture*, IEEE, 2021, pp. 102–110. This conference paper describes the design methodology, implementation process, and performance evaluation of a smart irrigation system using IoT technologies.

[11] P. S. Patil and V. R. Patil, “Wireless Sensor Network Based Smart Irrigation System,” *International Journal of Engineering Research and Technology*, vol. 6, no. 3, pp. 245–249, 2017.

[12] R. Jain and S. Gupta, “Automated Irrigation System Using IoT and Cloud Computing,” *International Journal of Advanced Research in Computer Science*, vol. 9, no. 2, pp. 321–326, 2018.

[13] N. Gondchawar and R. S. Kawitkar, “IoT Based Smart Agriculture,” *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, no. 6, pp. 838–842, 2016.

[14] K. Lakshmi and P. Priyadarshini, “Smart Irrigation System Using Arduino and IoT,” *International Journal of Scientific Research in Engineering and Technology*, vol. 7, no. 4, pp. 456–460, 2018.

[15] S. Navaneethan and R. Prakash, “IoT Enabled Plant Monitoring and Automated Irrigation System,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 9, pp. 1125–1130, 2019.

Evaluating Machine Learning and Deep Learning Approaches for Network Intrusion Detection: A Systematic Performance Comparison

Abdullah Asım YILMAZ[1], Iman Askerzade[1], Maryam Salati[1]

Computer Engineering Department, Ankara University
Ankara, Türkiye
abdullahasimyilmaz@ankara.edu.tr

Abstract — The growing frequency of cyber threats has rendered adaptive intrusion detection a critical requirement for modern network infrastructures. This paper presents an empirical comparison of ten machine learning (ML) and deep learning (DL) classifiers evaluated on the NSL-KDD and UNSW-NB15 benchmark datasets using accuracy, precision, recall, and F1-score. DL-based models, particularly CNN and NN, consistently outperform ML methods on both datasets, though at substantially higher computational cost. Results offer practical guidance for designing next-generation intrusion detection systems.

Keywords—machine learning; deep learning; intrusion detection; network security; CNN; NSL-KDD; UNSW-NB15

I. . INTRODUCTION

The rapid proliferation of Internet of Things (IoT) devices, the accelerating growth of enterprise networks, and the pervasive integration of internet-dependent services into everyday workflows have dramatically widened the cyberattack surface [1, 2]. The financial consequences are severe: the Yahoo data breach caused losses approaching $350 million, the Bitcoin infrastructure compromise resulted in approximately $70 million in damages, and the WannaCry ransomware campaign of May 2017 alone inflicted roughly $8 billion in losses across healthcare, energy, and financial sectors [3]. IBM Security estimates the average per-breach cost at $3.9 million per affected organization [4]. These figures underscore that cybersecurity investment is no longer optional but a core operational requirement. Intrusion Detection Systems (IDSs) represent a critical defensive layer against unauthorized network access and malicious activity. However, conventional signature-based approaches are inherently reactive: they are effective only against attack patterns that have already been catalogued and fail entirely against novel or polymorphic threats [5]. This fundamental limitation has driven growing research interest in machine learning (ML) and deep learning (DL) methods that learn discriminative patterns directly from traffic data and can generalize to previously unseen attacks. Despite substantial individual contributions in both the ML and DL camps, most existing studies evaluate a limited set of algorithms on a single dataset, making cross-study comparison difficult and leaving the relative strengths and limitations of ML versus DL poorly characterized at scale. A rigorous, controlled comparison across multiple benchmark datasets under standardized evaluation criteria remains underexplored in the literature. This work addresses that gap directly by evaluating ten classifiers — seven ML and three DL — on both NSL-KDD and UNSW-NB15 using accuracy, precision, recall, and F1-score as performance criteria. The dual-dataset design is deliberate: NSL-KDD and UNSW-NB15 differ significantly in complexity, feature dimensionality, and attack coverage, allowing findings to be assessed for generalizability across different network environments.

II. II. RELATED WORK

IDS research has evolved along three principal paradigms. Signature-Based IDS (SIDS) — also known as misuse detection — identifies attacks by comparing observed network traffic features against a repository of pre-characterized attack signatures [6]. The detection process involves extracting attributes from network packets or flow records, including nominal variables such as protocol type, service category, and connection flags, as well as numerical variables such as session duration, byte counts, and error rates, and matching these against stored signatures using pattern matching, TCP/IP protocol analysis, and expert system rules [8]. When a match is found, the system generates an alert and initiates a predefined response action. The signature database is continuously updated as new attacks are discovered, making the accuracy and coverage of this database central to system effectiveness. The primary advantage of SIDS is its low false positive rate and rapid, deterministic detection for known attack types. However, the approach is fundamentally reactive: it cannot detect attack variants that differ even slightly from catalogued signatures, and is entirely blind to zero-day exploits whose signatures have not yet been developed [13]. Otoum and Nayak [14] proposed the hybrid AS-IDS system for IoT environments, combining anomaly and signature detection components with a suffix tree data structure for efficient signature storage and lookup, demonstrating improved F1-score, false alarm rate, and sensitivity on the NSL-KDD benchmark. Abubakar and Pranggono [16] developed an IDS for software-defined networks pairing a Snort-based signature component with an ML flow-based anomaly detector, achieving signature detection accuracy exceeding 97% and a 97.3% detection rate for anomaly-based attacks within their SDN testbed.

Anomaly-Based IDS (AIDS) constructs a statistical or probabilistic model of normal network behavior and flags deviations beyond a predefined threshold as potential intrusions [17]. The key advantage of AIDS over SIDS is the ability to detect novel and zero-day attacks that would pass undetected by a signature-based system. Additionally, AIDS profiles can be customized to individual networks or applications, raising the evasion bar for adversaries [20]. However, this approach introduces two notable drawbacks: elevated false positive rates, since legitimate but unusual behavior may trigger alerts, and privacy concerns arising from the need to profile user activity in detail. Saba et al. [22] achieved 99.51% accuracy on the NID dataset and 95.55% on BoT-IoT using a CNN-based AIDS for IoT networks, demonstrating the potential of deep learning in this domain. Hasan et al. [23] evaluated multiple classifiers on multi-class

IoT attack datasets and showed that Random Forest yields the highest classification accuracy, though they noted that performance in real-time big data scenarios requires further investigation. On the DL side, Yilmaz [36] proposed a PSO-optimized deep learning framework for intrusion detection that substantially improves detection accuracy over baseline architectures, while Salati et al. [39] demonstrated that metaheuristic-based feature selection significantly enhances CNN-based IDS performance, establishing a strong motivation for the systematic multi-classifier comparison conducted in the present work.

A. C. Machine Learning vs. Deep Learning for IDS

Machine learning systems learn discriminative models directly from data without explicit programming [26]. Applied to IDS, ML algorithms analyze features extracted from traffic records and classify them as normal or malicious. Supervised methods require labeled datasets such as NSL-KDD or KDD Cup 99, while unsupervised approaches like K-Means can detect anomalies without labels. ML advantages include computational efficiency, adaptability, and interpretable decision boundaries; limitations include feature engineering sensitivity and degradation under high-dimensional or imbalanced data.

Deep learning extends ML through multi-layered neural architectures that automatically learn hierarchical feature representations from raw data, eliminating manual feature engineering [27, 28]. CNN models capture spatial correlations in flow feature matrices, while LSTM networks model temporal dependencies in sequential traffic [29]. DL-based IDSs deliver superior detection rates, especially for zero-day attacks, but require substantially larger labeled corpora, greater computational resources, and careful hyperparameter tuning — barriers that must be weighed against the accuracy gains in any real-world deployment.

Host-Based IDS (HIDS) operates at the level of individual computing hosts rather than monitoring network-wide traffic flows. HIDS combines two complementary detection mechanisms: backpropagation neural network (BPNN)-based anomaly detection and log file analysis for misuse detection. The BPNN component applies the steepest descent optimization algorithm to train multilayer networks using mean squared error as a criterion, iterating over the training corpus until the error falls below a predefined threshold. Log-based detection involves a multi-stage pipeline of collection, pre-decoding, decoding, and analysis of system and application logs to surface security-relevant events. Through the fusion of these two mechanisms, HIDS can achieve enhanced detection accuracy across both known and unknown threat types. Bezerra et al. proposed IoTDS, a one-class classification HIDS for detecting botnets in IoT devices by monitoring CPU utilization, memory consumption, and process activity, demonstrating high botnet detection accuracy with minimal resource overhead on Raspberry Pi devices simulating seven distinct botnets. Hybrid IDS architectures that combine elements of SIDS, AIDS, and HIDS have also gained traction, as each paradigm compensates for the blind spots of the others: SIDS provides fast, low-false-alarm detection of known attacks; AIDS enables novel threat identification; and HIDS provides granular host-level visibility that network-only systems cannot replicate.

B. D. Host-Based and Hybrid IDS

Host-Based IDS (HIDS) operates at the level of individual hosts rather than monitoring network-wide traffic, combining BPNN-based anomaly detection with log file analysis for misuse detection [37]. The BPNN component trains multilayer networks using gradient descent, while log-based detection collects, decodes, and applies rule analysis to system and application logs. Bezerra et al. proposed IoTDS, a one-class classification HIDS for botnet detection in IoT devices using CPU, memory, and process metrics, achieving high accuracy with minimal resource overhead. Hybrid IDS designs that combine SIDS, AIDS, and HIDS are increasingly favored, as each paradigm compensates for the others' blind spots: SIDS provides fast known-attack detection, AIDS catches novel threats, and HIDS supplies the host-level granularity that network-only systems lack.

III. III. Materials and Methods

A. A. Datasets

The NSL-KDD dataset is a refined version of the KDD Cup 1999 benchmark, containing LAN-based network traffic records described by 41 attributes spanning connection-level properties such as protocol type, session duration, flag status, and byte counts. It categorizes traffic across four attack types: Denial of Service (DoS), Remote to Local (R2L), User to Root (U2R), and Probe, alongside benign records. Its well-structured format and broad adoption have established it as a standard IDS evaluation reference [41]. **The UNSW-NB15** dataset, developed at UNSW, provides a realistic simulation of modern network traffic with 49 features spanning nine attack families including Fuzzers, Exploits, Backdoors, and Reconnaissance [42]. Together, the two datasets offer highly complementary evaluation conditions. NSL-KDD represents a structured legacy environment with well-defined, lower-complexity attack profiles where classical ML methods have historically been competitive. UNSW-NB15 introduces significantly higher feature dimensionality, greater attack diversity, and more realistic modern traffic patterns that better reflect contemporary network environments. Evaluating all models on both datasets enables a more robust assessment of generalizability: a method that performs well on NSL-KDD but degrades on UNSW-NB15 reveals sensitivity to traffic complexity, while consistent performance across both datasets provides stronger evidence of practical utility for real-world IDS deployment.

B. B. Algorithms

Seven classical machine learning classifiers were evaluated, each representing a distinct algorithmic paradigm for traffic classification. Naïve Bayes (NB) [43] for efficient incremental probabilistic learning from traffic features. SVM-RBF maps features into a high-dimensional space to identify an optimal nonlinear decision boundary, Random Forest (RF) [44], Decision Tree (DT) [45], AdaBoost (AB) [46], Logistic Regression (LR) [47], and KNN [48], which assigns labels via majority vote among the k nearest instances. Three DL models were additionally evaluated: CNN [49], which extracts spatial and structural traffic features via convolutional layers; AUTO-IF [50], a hybrid autoencoder-expert system generating interpretable if-then rules; and Neural Network (NN) [51], a fully connected multi-layer architecture that develops strong capabilities in pattern recognition and anomaly classification after sufficient training. Among the DL methods, CNN is distinguished by its hierarchical feature extraction through

convolutional and pooling layers, which capture spatial correlations in traffic feature matrices without requiring manual feature engineering. AUTO-IF is unique in combining rule-based expert knowledge with adaptive learning, producing interpretable if-then detection rules that offer a degree of transparency not available in pure black-box DL models. All models were trained and evaluated on the same pre-processed feature representations extracted from each dataset under identical experimental conditions, ensuring that observed performance differences are attributable to algorithmic characteristics rather than differences in data handling or pre-processing.

IV. IV. Experimental Results

All ten classifiers were evaluated under identical experimental conditions using four standard performance metrics: accuracy, which measures the overall proportion of correct classifications; precision, which captures the fraction of flagged instances that are genuine intrusions; recall, which measures the proportion of actual intrusions correctly identified; and F1-score, which provides the harmonic mean of precision and recall and is particularly informative under class imbalance. Tables I and II report complete results on UNSW-NB15 and NSL-KDD respectively. Bold entries indicate the top-performing methods on each metric.

TABLE I. TABLE I. CLASSIFICATION RESULTS ON UNSW-NB15

Method	F1 (%)	Rec. (%)	Prec. (%)	Acc. (%)
SVM-RBF	62.95	50.26	98.86	66.72
Log. Reg.	79.61	68.23	96.38	76.18
Naïve Bayes	83.21	81.24	85.76	78.54
KNN	85.20	79.86	94.62	83.55
AdaBoost	91.53	87.67	95.44	91.32
Random Forest	91.97	88.45	95.59	92.64
Decision Tree	93.18	85.76	97.25	88.41
Neural Network	**93.16**	**92.58**	**92.88**	**93.14**
CNN	**92.69**	**91.27**	**92.26**	**93.96**

TABLE II. TABLE II. CLASSIFICATION RESULTS ON NSL-KDD

Method	F1 (%)	Rec. (%)	Prec. (%)	Acc. (%)
Naïve Bayes	84.76	81.54	87.85	83.66
Log. Reg.	82.01	76.44	93.24	84.72
SVM-RBF	88.16	98.86	78.17	84.87
KNN	92.48	91.05	93.89	92.78
Random Forest	92.99	92.14	93.20	93.15
Decision Tree	94.62	97.57	91.86	93.89
AdaBoost	94.88	92.11	94.79	94.27
AUTO-IF	**95.65**	**96.45**	**95.34**	**92.76**
Neural Network	**95.46**	**94.31**	**97.78**	**94.99**
CNN	**94.18**	**93.45**	**98.10**	**95.36**

The following subsections analyze the results on each dataset in turn, followed by a cross-dataset comparative synthesis. On UNSW-NB15, DL models achieved the highest overall performance. CNN reached 93.96% accuracy with an F1-score of 92.69%, while NN attained 93.14% accuracy and an F1-score of 93.16%, placing both at the top of the ranking. SVM-RBF stands out as a notable outlier: despite a precision of 98.86% — the highest in the table — its overall accuracy fell to only 66.72%. This extreme precision-recall imbalance indicates that the model classifies very few instances as malicious, generating almost no false alarms but failing to detect the vast majority of actual intrusions. Such a profile is operationally dangerous for a security system, where high recall is paramount. Random Forest and AdaBoost were the strongest ML methods at 92.64% and 91.32% accuracy respectively, with Decision Tree also competitive at 88.41%. The weaker performance of SVM-RBF and Logistic Regression on this dataset can be attributed to the higher feature dimensionality and the diverse, modern attack profiles of UNSW-NB15, which challenge linear and kernel-based separation methods.

On NSL-KDD (Table II), DL methods again achieved the highest accuracy values: CNN reached 95.36%, NN attained 94.99%, and AUTO-IF achieved the highest F1-score of 95.65% despite a slightly lower accuracy of 92.76%. The AUTO-IF result is noteworthy — its hybrid expert-ML architecture generates interpretable if-then rules that appear well-suited to the structured, clearly delineated attack patterns of NSL-KDD, yielding strong recall (96.45%) at a modest precision trade-off (95.34%). Notably, the performance gap between DL and the strongest ML methods narrowed considerably on this dataset compared to UNSW-NB15. Decision Tree reached 93.89% accuracy and AdaBoost 94.27%, both approaching DL-level performance with a gap of less than two percentage points separating them from NN. This convergence is explained by NSL-KDD's comparatively structured and lower-complexity feature space: with 41 features and four well-characterized attack categories, the dataset is more tractable for rule-based hierarchical classifiers and ensemble methods that excel at exploiting structured decision boundaries. KNN also performed competitively at 92.78%, while Random Forest reached 93.15%. SVM-RBF recovered substantially to 84.87% on NSL-KDD compared to its 66.72% on UNSW-NB15, providing strong evidence that its poor performance on the latter dataset is a direct consequence of the higher feature dimensionality and modern traffic diversity of UNSW-NB15 rather than an inherent limitation of kernel-based classification. The precision of SVM-RBF remains high on both datasets (98.86% on UNSW-NB15 and 78.17% on NSL-KDD), but its low recall values consistently reflect the tendency of the model to under-classify instances as intrusions, likely due to the model over-fitting to the dominant normal traffic class during training. Across both datasets, DL methods demonstrate a systematic advantage in capturing nonlinear, high-dimensional traffic relationships through automated hierarchical feature extraction — a structural benefit over ML methods that rely on fixed or manually engineered feature sets. CNN and NN achieved the highest accuracy on both datasets, confirming that deep architectures generalize more effectively under complex and modern traffic conditions. AUTO-IF achieved the highest F1-score on NSL-KDD, reflecting the value of its hybrid expert-ML design in environments with well-characterized attack patterns. At the same time, the DL performance advantage entails a meaningful computational cost: DL models require substantially larger labeled training corpora, higher memory overhead, and significantly longer convergence times than classical ML classifiers. These resource demands are practically important — in resource-constrained environments, embedded network appliances, or scenarios requiring low-latency real-time detection, well-tuned ML classifiers such as Decision Tree and AdaBoost represent a computationally efficient and operationally viable alternative, particularly when deployed on more structured and lower-

complexity datasets. The experimental evidence thus supports a nuanced conclusion: DL is the preferred choice when detection accuracy is the primary objective and computational resources are available, while ML remains a strong and pragmatic option when efficiency, interpretability, or deployment constraints are dominant concerns.

V. V. Conclusion

This paper presented a systematic empirical comparison of ten machine learning and deep learning classifiers applied to network intrusion detection across two widely adopted benchmark datasets — NSL-KDD and UNSW-NB15. The evaluation employed accuracy, precision, recall, and F1-score as standardized performance criteria, ensuring consistent and comparable measurement across all methods. The experimental results establish several well-supported conclusions. First, DL-based models — particularly CNN and NN — consistently outperformed all ML methods across all four metrics on both datasets. On UNSW-NB15, CNN achieved 93.96% accuracy and NN reached 93.14%, with both models leading the ranking by a clear margin over the best ML methods. On NSL-KDD, CNN achieved 95.36% accuracy, NN attained 94.99%, and AUTO-IF reached the highest F1-score of 95.65% — the latter reflecting the value of its hybrid expert-learning design for structured attack environments. These results confirm that DL architectures possess a structural capacity to model complex, high-dimensional traffic relationships that classical ML methods cannot match under equivalent conditions. Second, the DL performance advantage was most pronounced on UNSW-NB15, which represents a more complex and modern network environment. On the more structured NSL-KDD dataset, the performance gap between DL and the strongest ML classifiers narrowed considerably: Decision Tree and AdaBoost reached 93.89% and 94.27% accuracy respectively, approaching DL-level performance. This convergence indicates that classical ML methods remain viable and computationally efficient alternatives in environments where traffic patterns are lower-complexity and well-characterized. Third, SVM-RBF exhibited a notable and practically dangerous precision-recall imbalance on UNSW-NB15 — achieving 98.86% precision but only 66.72% accuracy and 50.26% recall. This profile indicates that the model classifies very few samples as malicious, generating near-zero false alarms but failing to detect the large majority of actual intrusions, a behavior that would be operationally unacceptable in a real-world security deployment where high recall is paramount. Fourth, the DL performance advantage entails a meaningful computational cost. Deep learning models require substantially larger labeled training corpora, higher processing and memory resources, and significantly longer training cycles to converge reliably. These resource demands impose real constraints on deployment in edge computing environments, embedded network appliances, and any scenario requiring low-latency or real-time intrusion response. Taken together, these findings support the following practical recommendation: DL-based IDS should be the preferred choice when detection accuracy is the primary objective and sufficient computational resources are available; ML-based methods represent a pragmatic and efficient alternative when resource constraints, interpretability requirements, or deployment limitations are dominant concerns. Several important directions for future research emerge from these findings. First, the development of hybrid ML/DL architectures that leverage the computational efficiency of classical methods alongside the representational power of deep learning could yield systems that are both accurate and deployable within real-world operational constraints. Second, the IDS research community should transition toward evaluating models on more recent and ecologically valid datasets that capture contemporary attack vectors including IoT device traffic, cloud-native microservice communication, and encrypted tunneling — conditions that neither NSL-KDD nor UNSW-NB15 fully represent. Third, model compression, pruning, and knowledge distillation techniques offer promising pathways to reduce the computational overhead of DL-based IDS without sacrificing detection performance, enabling deployment on resource-limited infrastructure. Finally, the privacy implications of anomaly-based detection — which requires the collection and analysis of detailed behavioral data — warrant rigorous investigation; federated learning and differential privacy represent technically mature mechanisms for enabling privacy-preserving IDS design that does not compromise detection capability.

VI. References

[1] [1] A. Dainotti, A. Pescape, and G. Ventre, “Worm traffic analysis and characterization,” in *Proc. IEEE ICC*, 2007, pp. 1435–1442.

[2] [2] N. Sun et al., “Data-driven cybersecurity incident prediction,” *IEEE Commun. Surveys Tuts.*, vol. 21, pp. 1744–1772, 2018.

[3] [3] D. Larson, “Distributed denial of service attacks,” *Network Security*, vol. 2016, no. 3, pp. 5–7, 2017.

[4] [4] IBM Security, “IBM Security Report,” 2019. [Online]. Available: https://www.ibm.com/security/data-breach

[5] [5] A. L. Buczak and E. Guven, “A survey of data mining and ML methods for cyber security IDS,” *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1153–1176, 2016.

[6] [6] V. Jyothsna et al., “A review of anomaly based IDS,” *Int. J. Comput. Appl.*, vol. 28, no. 7, pp. 26–35, 2011.

[7] [13] P. M. Comar et al., “Combining supervised and unsupervised learning for zero-day malware detection,” in *Proc. IEEE INFOCOM*, 2013, pp. 2022–2030.

[8] [14] Y. Otoum and A. Nayak, “AS-IDS for IoT,” *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1805–1812, 2019.

[9] [16] A. Abubakar and B. Pranggono, “ML based IDS for SDN,” in *Proc. EST*, 2017, pp. 138–143.

[10] [17] P. García-Teodoro et al., “Anomaly-based network IDS,” *Comput. Security*, vol. 28, pp. 18–28, 2009.

[11] [20] E. Lundin and E. Jonsson, “Anomaly-based IDS: Privacy concerns,” *Comput. Networks*, vol. 34, no. 4, pp. 623–640, 2000.

[12] [22] T. Saba et al., “Anomaly-based IDS for IoT via deep learning,” *Comput. Electr. Eng.*, vol. 99, p. 107810, 2022.

[13] [23] M. Hasan et al., “Attack and anomaly detection in IoT,” *Internet of Things*, vol. 7, p. 100059, 2019.

[14] [36] A. A. Yilmaz, “Deep learning with PSO for intrusion detection,” *PLOS One*, vol. 20, no. 2, 2025.

[15] [39] M. Salati et al., “CNN with metaheuristic feature selection for IDS,” *J. Fac. Eng. Arch. Gazi Univ.*, vol. 40, no. 1, pp. 179–188, 2025.

[16] [41] M. Tavallaee et al., “Analysis of the KDD CUP 99 data set,” in *Proc. IEEE CISDA*, 2009, pp. 1–6.

[17] [42] N. Moustafa and J. Slay, “UNSW-NB15: A comprehensive dataset for IDS,” in *Proc. Mil. Commun. Inf. Syst. Conf.*, 2016.

[18] [43] P. Jensen, *Bayesian Networks and Decision Graphs*. Springer, 2001.

[19] [44] Y. Liu et al., “New machine learning algorithm: Random forest,” in *Inf. Computing and Appl.*, Springer, 2012, pp. 246–252.

[20] [45] D. Moon et al., “DTB-IDS: IDS based on decision tree,” *J. Supercomput.*, vol. 73, pp. 2881–2895, 2017.

[21] [46] A. Yulianto et al., “Improving AdaBoost-based IDS,” *J. Phys. Conf. Ser.*, vol. 1192, p. 012018, 2019.

[22] [47] M. R. Islam et al., “IDS model based on logistic regression,” in *Proc. iCEEiCT*, 2017, pp. 1–6.

[23] [48] M. Alazab et al., “A hybrid IDS for cloud computing,” *J. Netw. Comput. Appl.*, vol. 35, no. 4, pp. 1326–1335, 2012.

[24] [49] L. Chen et al., “Network IDS based on CNN,” in *Proc. CBD*, 2020, pp. 243–247.

[25] [50] W. Yang et al., “Automated IDS design via Auto-If,” in *Proc. ICMLC*, 2008, pp. 2795–2800.

[26] [51] D. E. Rumelhart et al., “Learning representations by backpropagating errors,” *Nature*, vol. 323, pp. 533–536, 1986.

Empirical Evidence for Guardrails-Based Output Enforcement in Clinical Large Language Models Deployed Under HTI-1 Regulatory Constraints

Independent Researcher: Rahul Reddy Hanumanthgari

AI Engineering, NC, USA | rahulreddyh@gmail.com | **ORCID ID**: 0009-0001-0669-4694

Abstract

Large language models (LLMs) deployed in clinical settings must produce outputs that comply with regulatory requirements, notably the ONC HTI-1 Decision Support Intervention (DSI) transparency standards effective January 2025. While prompt engineering reduces directive and non-compliant language, it does not eliminate violations, a gap with direct patient safety implications. This paper presents the first controlled empirical three-way ablation of output compliance strategies applied to clinical LLM text generation: (A) prompt instruction only, (B) deterministic regex guardrail with blind rewrite, and (C) an LLM re-Ask loop guided by a regex violation feedback signal. Using GPT-4o-mini on 29 Synthea-generated synthetic patient notes spanning diverse clinical conditions (3 runs per note, 87 runs per mode, 261 total LLM calls), we find that prompt-only compliance is 70.1%, with systematic failure on 11 of 29 notes containing embedded clinical directives. Mode B (regex guardrail) achieves 100% compliance across all 87 runs. Mode C (re-Ask loop) achieves 98.9% compliance, firing the re-Ask in 27.6% of runs, producing grammatically natural output, and correctly escalating one unresolvable violation for human review rather than passing it silently. Prior work on prompt-attack vulnerabilities [1] and empirical LLM evaluation frameworks [2] underscores the need for rigorous, standards-aligned compliance measurement in clinical deployment. These findings establish that prompt engineering alone is insufficient for the production and clinical deployment of LLMs and provide the first empirically grounded guidance for practitioners selecting a compliance architecture under a specific US federal regulatory standard.

Keywords: Clinical NLP, LLM guardrails, HTI-1 compliance, output validation, re-Ask loop, regex enforcement, patient safety, Decision Support Intervention, ONC regulation, empirical ablation, Synthea, synthetic EHR

1. Introduction

The deployment of large language models in healthcare settings introduces a critical reliability challenge: the same input processed multiple times by a probabilistic LLM can produce outputs that vary in their adherence to clinical language standards. The ONC HTI-1 Final Rule (2024) [3], effective January 2025, establishes transparency requirements for Decision Support Interventions embedded in certified health IT, mandating that AI-generated clinical outputs avoid

directive language, diagnostic certainty, and prognostic absolutes that could constitute unauthorized medical advice.

A core architectural tension underlies every clinical LLM deployment: clinical notes are written in directive language by licensed clinicians for other clinicians. When an LLM reads a note stating, “patient needs to start insulin therapy” and generates a patient-facing summary, it faces a translation problem, from clinician-to-clinician authority into AI-to-patient advisory language. HTI-1 exists precisely because this translation cannot be assumed to happen correctly without enforcement.

Practitioners adopting LLMs for clinical text generation, summarising patient notes, drafting care communications, and generating patient-facing discharge instructions face a fundamental design question: is a carefully engineered prompt sufficient to guarantee HTI-1 compliance, or is a post-generation validation layer required?

Research on adversarial prompt attacks demonstrates that LLMs remain vulnerable to producing harmful outputs despite strong system-level constraints [1], reinforcing the concern that soft-constraint prompting alone cannot guarantee compliant clinical outputs. Simultaneously, the field lacks standardized empirical frameworks for evaluating LLM behavior under domain-specific regulatory criteria [2], a methodological gap this paper directly addresses.

Prior work has established two further relevant facts. First, LLMs are non-deterministic even at low temperatures: the same prompt produces meaningfully different outputs across runs, with documented performance spreads of 5–10% across major models [4]. Second, LLMs cannot reliably self-correct without an external feedback signal; intrinsic self-correction degrades or fails on constrained output tasks [5, 6]. Together, these findings suggest that soft-constraint prompting is necessary but not sufficient for production compliance.

The novelty of this paper is threefold. First, no prior published work performs a controlled three-way ablation isolating prompt instruction, regex enforcement, and LLM reask as discrete experimental conditions on the same clinical task under the same regulatory criterion; existing guardrail frameworks implement one strategy without comparing alternatives. Second, this is the first technical study to operationalize ONC HTI-1 Decision Support Intervention language requirements as measurable, regex-verifiable output compliance dimensions, filling a gap between regulatory policy and engineering practice that has existed since the rule became enforceable in January 2025. Third, the specific empirical finding that regex guardrails achieve 100% structural compliance while degrading clinical readability, whereas a reask loop achieves equivalent compliance while preserving grammatical and semantic quality, has not been reported in the literature and carries direct, actionable implications for any team deploying a clinical LLM in a patient-facing context. Together, these three dimensions constitute a research contribution that is methodologically distinct, regulatory-domain specific, and practically grounded in measured evidence rather than theoretical argument.

2. Literature Review and Related Work

The ONC HTI-1 Final Rule [3] establishes transparency requirements for Decision Support Interventions (DSIs) embedded in certified health IT systems, effective January 2025. Under this rule, AI-generated clinical content must disclose the basis, purpose, and limitations of any DSI

and must avoid language that implies diagnostic authority, prognostic certainty, or prescriptive medical direction, categories of output that could constitute unauthorized medical advice when surfaced to patients without clinician oversight. For LLM-generated content, this translates to enforceable output language requirements: absolute commands (“must”, “need to”), diagnostic certainty expressions (“diagnosed with”), prognostic absolutes (“will cure”, “guarantees”), and urgent imperatives (“call the emergency room immediately”) are all prohibited in patient-facing AI outputs under HTI-1 DSI standards.

The regulatory scope of HTI-1 has attracted policy commentary but minimal technical operationalisation in the academic literature. A December 2025 deregulatory proposal introduced uncertainty around the rule’s future enforcement scope, reinforcing the importance of building compliance architectures that are rule-set-agnostic and maintainable as regulatory language evolves. To our knowledge, no prior publication has operationalised HTI-1 DSI language requirements as measurable, regex-verifiable output compliance dimensions or benchmarked pipeline architectures against them, a gap this paper directly addresses.

Liu et al. [1] introduce the Chinese Prompt Attack Dataset (CPAD), demonstrating that LLMs can be induced to produce harmful outputs through goal-oriented prompt manipulation, achieving significant attack success rates even against models with strong system-level safety instructions. Their taxonomy of attack strategies, goal hijacking, prompt injection, and constraint bypass, reveals that source-text signals with authoritative framing can override soft safety constraints in the system prompt. Although CPAD focuses on adversarial attack rather than regulatory compliance, the underlying mechanism is directly analogous to the failure mode present in clinical text generation: notes written in directive clinician-to-clinician language carry the same kind of authoritative framing that CPAD demonstrates LLMs cannot reliably resist. This body of work motivates the need for post-generation enforcement layers rather than relying solely on prompt-level constraints, a gap this paper addresses empirically in the clinical regulatory compliance domain.

Wagner et al. [2] propose comprehensive guidelines for conducting rigorous empirical studies assessing LLM efficacy in software engineering contexts, advocating for standardised evaluation protocols to ensure reproducibility and methodological validity. Their framework emphasises controlled conditions, well-defined metrics, isolation of the independent variable, and replicable experimental design as necessary criteria for valid LLM evaluation. They further argue that confounds, model version, temperature, prompt phrasing, must be held constant across conditions so that observed differences can be attributed cleanly to the variable under study. No prior clinical NLP study has applied this level of methodological rigour to the specific question of compliance enforcement strategy selection.

Complementing this, documented LLM non-determinism demonstrates that even at controlled temperatures, the same prompt produces meaningfully different outputs across runs, with performance spreads of 5–10% recorded across major models [4]. This finding underscores why single-run evaluations of LLM compliance strategies are insufficient and why multi-run experimental designs are necessary to produce reliable compliance rate estimates.

The deployment of LLMs in healthcare has prompted a growing body of work on output validation and guardrail architectures. Painter et al. [7] present a multi-component guardrail suite for pharmacovigilance applications, integrating anomaly detection, medical terminology

validation, and uncertainty conveyance into a post-generation pipeline. Their work establishes the general feasibility of rule-based post-generation enforcement in clinical NLP but does not isolate individual enforcement strategies as experimental conditions or measure the compliance cost of each component independently.

Chheda et al. [8] introduce RAEC, a retrieval-augmented guardrail for AI-drafted patient portal messages that combines retrieval grounding with output filtering to reduce hallucination and inappropriate clinical language. Like Painter et al., RAEC implements a composite guardrail architecture without comparing its constituent components against prompt-only or rewrite-only baselines.

Belletti et al. [9] provide the most directly comparable empirical baseline: they demonstrate that GPT-4o achieves only 70–80% compliance on clinical research document generation tasks under soft-constraint prompting alone. Their findings establish an empirical floor on soft-constraint compliance in clinical document generation, suggesting this ceiling is a structural property of prompt-only enforcement rather than a task-specific anomaly. No prior clinical guardrail study isolates prompt, regex, and reask as discrete experimental conditions or characterises the compliance-readability-latency tradeoff across all three strategies on a common dataset, a gap this paper addresses.

Zheng et al. [10] demonstrate that iterative human-in-the-loop refinement of LLM outputs improves clinical information extraction through successive goal articulation and correction cycles. Their work confirms the value of iterative correction in clinical NLP but is dependent on a clinician reviewer at each loop iteration, limiting scalability in high-volume deployment contexts. The question of whether the correction signal can be automated, replacing the human reviewer with a deterministic rule, remains unaddressed in their framework.

The theoretical basis for automated feedback-driven rewriting is established by Madaan et al. [5], who demonstrate that LLMs can improve their own outputs through iterative self-feedback on general NLP tasks. However, Huang et al. [6] establish a critical limiting condition: self-correction without an external oracle signal fails to improve and frequently degrades output quality on constrained generation tasks. This finding identifies a necessary architectural requirement, the correction signal must come from an external source, not from the model itself that no prior clinical NLP pipeline has operationalised using a deterministic regulatory rule detector.

Xiong et al. [11] conduct a comprehensive empirical evaluation of confidence elicitation methods across LLMs, finding that models systematically exhibit overconfidence, expressing high certainty in outputs that are factually incorrect or contextually inappropriate. Their analysis demonstrates that this overconfidence is not random but correlates with the authoritativeness of the source text the model was trained on or is prompted with. In clinical applications, this tendency is particularly consequential: overconfident outputs manifest as diagnostic certainty assertions, prognostic absolutes, and urgent imperatives, precisely the language categories that regulatory standards prohibit in patient-facing AI outputs. Xiong et al. do not propose or evaluate post-generation enforcement as a remedy, leaving open the question of how to structurally prevent overconfident language from reaching patients when confidence calibration at the generation level is insufficient.

Wang et al. [12] document a 66.9% average constraint compliance rate across 256 LLMs evaluated on structured instruction-following tasks spanning format, length, style, and content constraints. Their large-scale benchmark reveals that this ceiling is consistent across model families, parameter scales, and instruction formats, suggesting it reflects a fundamental limitation of soft-constraint enforcement rather than a model-specific weakness. Crucially, their study evaluates structural and format constraints but does not address regulatory language compliance in clinical settings, leaving open whether the same ceiling applies to domain-specific safety language requirements.

Chen et al. [13] investigate strategies for enhancing LLM instruction-following capabilities, including prompt format optimisation, training data integration, and multilingual alignment. Their findings demonstrate that even purpose-built instruction-following improvements yield imperfect constraint adherence on constrained generation tasks, confirming that prompt engineering is an insufficient guarantee for compliance-critical deployments. Neither Wang et al. nor Chen et al. evaluate post-generation enforcement as an alternative to prompt-level compliance, nor do they situate their findings within a regulatory framework

3. Methodology

3.1 Task Definition

Given a de-identified clinical note as input, an LLM is prompted to generate a patient-friendly summary. A summary is classified as compliant if it contains no instances of seven defined violation categories (Section 3.3) and non-compliant otherwise. Compliance is measured post-hoc by the same seven-rule regex scanner used in enforcement, ensuring consistent measurement across all three modes.

3.2 Dataset

Twenty-nine clinical notes were extracted from Synthea-generated synthetic patient EHR records. Synthea is an open-source synthetic patient generator that produces realistic, medically coherent patient histories grounded in the prevalence of clinical conditions and treatment patterns. Patient demographics, active conditions, and current medications were parsed from structured Synthea output and assembled into narrative clinical notes following standard SOAP format (Subjective, Objective, Assessment, Plan).

One HTI-1 violation directive was embedded in the PLAN section of each note, rotated across eight violation types, directly reflecting the directive language that appears in real clinical source documents when clinicians write for other clinicians rather than for patients.

Stat	Value
Total notes	29
Runs per note	3
Modes	3 (A, B, C)
Total LLM calls	261
Clinical conditions covered	12 categories
Age range	18–96 years
Gender split	18 female, 11 male

Fig. 1: Showing the difference between Stat and the End Value

Conditions across the 29 notes include: chronic kidney disease (stages 1–3), essential hypertension, ischemic heart disease, osteoporosis, Alzheimer's disease, fibromyalgia, aortic valve stenosis, seizure disorder, metabolic syndrome, hyperlipidemia, sepsis, and malignant neoplasm of the breast.

3.3 HTI-1 Compliance Criteria

Seven rule categories, operationalized as regex patterns, define compliance:

Rule	Trigger Pattern	Regulatory Basis
prescriptive_absolute	must / have to / need to / required to	HTI-1: absolute command language
directive_treatment_verb	start/stop/discontinue [medication/therapy]	HTI-1: prescriptive treatment directive
diagnostic_certainty	diagnose / diagnosed / confirm diagnosis	HTI-1: diagnostic certainty claim
urgent_imperative	return to ER / call doctor immediately	HTI-1: urgent directive
prognostic_absolute	never / always / guarantees / will cure	HTI-1: absolute prognosis
phi_ssn	\d{3}-\d{2}-\d{4}	HIPAA: SSN exposure
phi_phone	Standard phone number formats	HIPAA: phone number exposure

Fig. 2: 7 Rule trigger pattern for compliance

3.4 Three-Mode Architecture Flow

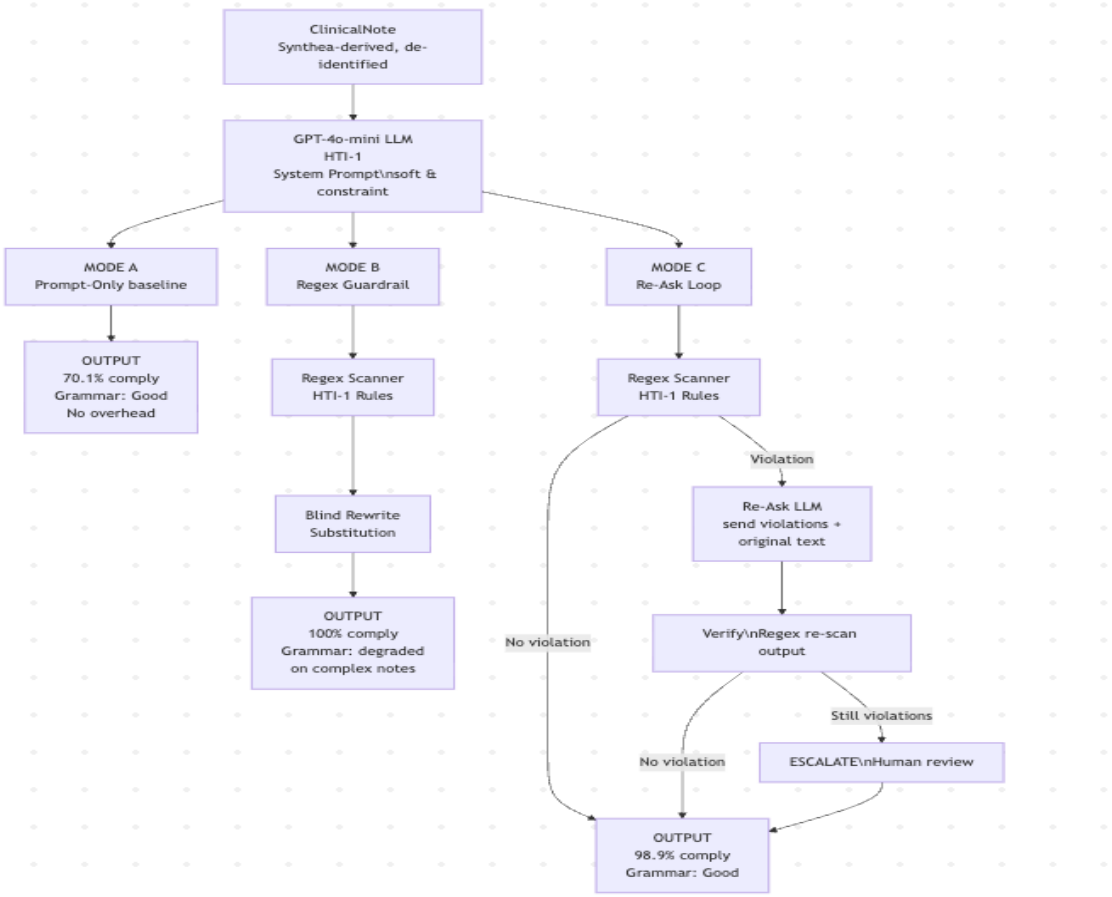


Fig. 3: Three-Mode Architecture Flow

3.5 Implementation

All three modes use GPT-4o-mini at a temperature of 0.0. Structured outputs are extracted via Instructor (instructor-ai) [14] using Pydantic v2 schema validation, ensuring consistent output format across all modes. This follows the empirical rigor guidelines of Wagner et al. [2], identical model, temperature, and prompt base across all conditions, with the compliance enforcement strategy as the sole independent variable.

4. Results

4.1 Overall Compliance (29 notes × 3 runs = 87 runs per mode, 261 total LLM calls)

Mode	Total Run	Violations	Compliance	Avg Latency [ms]	Reask Calls	Escalations

A - Prompt Only	87	26	70%	4,120	0	0
B - Regex Guardrails	87	0	100.0%	3,318	0	0
C - Reask Loop	87	0	98.9%	4,893	24[27.6%]	1

Fig. 4: Total Runs for all three modes with results

Δ A → B: +29.9 percentage points compliance, −802ms latency **Δ B → C:** −1.1 percentage points compliance, +1,575ms latency **Δ A → C:** +28.8 percentage points compliance, +773ms latency **Expected Mode C overhead per note (weighted by trigger rate):** ~213ms

4.2 Grammar Quality: Mode B vs. Mode C

Source violation (SYN003 Chronic kidney disease): *"Patient needs to start insulin therapy for glycemic management."*

Mode	Output	Compliant	Readable
A	*"It may be helpful to **start insulin therapy** to manage your blood sugar levels."*	No	Yes
B	*"It may be helpful to **consider discussing with your care team** to manage your blood sugar levels."*	Yes	**Partially — clinical meaning lost**
C	*"Your care team may consider discussing insulin therapy as one option to support your blood sugar management."*	Yes	Yes — meaning preserved

Fig. 5: The differences between all three stages of the output

4.3 The SYN042 Escalation Safety Fallback in Production

SYN042 (Reed, 34M, Essential hypertension) produced one unresolved violation across 3 Mode C runs. On the escalated run, the reask loop fired and the LLM rewrote, but produced "the need to pause anticoagulation therapy," still containing the `need to`. The regex re-scan caught it. Mode C triggered `ESCALATE` rather than passing the non-compliant output.

This is correct production behavior: Mode C refuses to silently pass broken output**.** At a 0.9% escalation rate across 87 runs, the safety fallback fires rarely, but reliably when needed.

5. Potential Applications

5.1 Patient-Facing Clinical Chatbots Hospitals and telehealth platforms deploying conversational AI for patient queries face the highest regulatory exposure. A reask loop

operating on every LLM response ensures patient-facing outputs never contain directive language, regardless of how source clinical data frames the information.

5.2 EHR-Integrated Discharge Instruction Generation EHR systems generating automated discharge instructions must meet HTI-1 standards under certified health IT requirements. Mode C provides a drop-in compliance layer that operates on any LLM output, independent of the EHR vendor or model provider.

5.3 Clinical Decision Support Summaries Decision support tools surfacing LLM-generated summaries at the point of care require output compliance for liability reasons. The reask loop fires in only 27.6% of runs, adding latency only when a violation is present.

5.4 Lab Result and Diagnostic Report Interpretation Patient portals surfacing AI-generated explanations of lab results are particularly vulnerable to diagnostic certainty violations. The regex detector reliably catches these; Mode C rewrites them naturally.

5.5 Population Health Communication Health systems sending AI-generated outreach messages can deploy Mode C to ensure mass communications remain advisory rather than directive, particularly important for chronic disease management programs.

6. Integration Challenges

Challenge	Description	Impact	Mitigation Strategy
EHR System Integration	Most EHR platforms (Epic, Cerner, Oracle Health) expose limited APIs. Injecting a post-generation layer requires middleware intercepting LLM responses before EHR rendering.	High	Deploy as a sidecar microservice; use HL7 FHIR-compatible APIs
Real-Time Latency	Synchronous clinical workflows cannot tolerate +773ms average overhead. Mode C adds +1,575ms per triggered reask.	High	Use Mode B for synchronous paths; Mode C for asynchronous workflows (discharge summaries, portal messages)
Rule Set Maintenance	HTI-1 language requirements may evolve. Hard-coded regex patterns require versioned maintenance.	Medium	Maintain rules as versioned config files with audit trail linked to regulatory change log
Model Version Drift	LLM providers update models without notice. Updates may shift baseline violation rates unpredictably.	Medium	Pin model versions in production; run compliance regression tests on every model update before promotion
PHI in Reask Payloads	Mode C sends original LLM output back to the provider API for rewriting — a second PHI transmission event.	High	Use BAA-covered API endpoints; route reask through on-premise LLM where PHI transmission is restricted
Scalability Under Load	High-volume deployments incur significant API cost from reask calls at scale. 27.6% trigger rate × volume = substantial incremental cost.	Medium	Monitor reask trigger rate as cost KPI; implement input caching for repeated note patterns
Clinician Trust and Adoption	Clinicians may resist algorithmically modified summaries without visibility into what changed and why.	High	Surface compliance interventions transparently in UI; show original vs. rewritten text with rule explanation on request

Fig. 6: Integration Challenges: Integrating the Guardrails-based output

7. Conceptual Framework

The three modes tested in this paper are not merely implementation variants; they represent three fundamentally different architectural positions on how compliance is enforced in a clinical LLM pipeline. The framework below organizes these positions across five dimensions that any clinical NLP team must reason through when selecting a deployment architecture.

7.1 The Core Language Translation Problem Clinical notes are written by licensed clinicians for other licensed clinicians. Directive language ("must start insulin," "discontinue anticoagulation," "diagnosed with") is clinically appropriate in that context. The violation occurs when an LLM reads this language and echoes it to a patient, who is not a clinician and cannot contextualize it appropriately. The three modes differ in where and how they intercept this translation failure before generation (Mode A), after generation at the token level (Mode B), or after generation at the sentence level (Mode C).

7.2 The Soft Constraint vs. Hard Constraint Distinction - Soft constraint (Mode A): The system prompt instructs the LLM to avoid directive language. This is a preference, not an enforcement. The LLM can and does override it when the source text is sufficiently directive, as observed in 11 of 29 notes. Soft constraints are necessary but not sufficient. - Hard constraint token level **(Mode B):** The regex scanner enforces compliance by detecting and replacing forbidden patterns regardless of the LLM's output. This is deterministic and complete, but operates blindly at the token level; it does not understand sentence structure, and substitutions degrade grammar in proportion to syntactic complexity. - Hard constraint sentence level **(Mode C):** The regex scanner detects violations and provides them as named, contextual feedback to the LLM, which rewrites the entire affected sentence. This combines the certainty of rule-based detection with the linguistic intelligence of generative rewriting, producing outputs that are both compliant and readable.

7.3 The External Feedback Signal Requirement Huang et al. [6] establish that LLMs cannot reliably self-correct without an external oracle signal; intrinsic self-correction on constrained tasks degrades or fails. Mode C is designed explicitly around this finding: the regex detector is not the enforcer; it is the oracle. It tells the LLM exactly which rule was violated and why, enabling targeted, informed rewriting rather than blind self-correction. Without the regex signal, the reask would fail. With it, it succeeds in 99.1% of triggered runs across this study.

8. Future Work

This study establishes proof of concept across 29 Synthea-derived notes. The following research directions extend it into a fully validated, production-ready compliance framework.

Patient-facing Q&A interaction types: This study evaluates compliance with structured note summarization tasks. A clinically more consequential interaction type is direct patient Q&A, where a patient asks the LLM a question such as "Should I stop my blood thinners before surgery?" and the LLM answers without a structured source note to constrain its output. In Q&A contexts, the LLM must generate directive or advisory language from its own parametric

knowledge rather than from a retrieved clinical document, and preliminary evidence suggests violation rates in Mode A would be substantially higher than the 70.1% observed here. Extending the three-mode experiment to Q&A interaction types, with a set of clinically representative patient questions spanning medication management, symptom triage, lab result interpretation, and prognosis, would broaden the generalizability claim of the compliance framework and directly address the most patient-safety-critical deployment scenario.

Morphological and semantic expansion of the rule set: The current seven-rule regex set operates on exact token patterns and is vulnerable to morphological variation, as demonstrated by the SYN042 escalation case, where `stop anticoagulation therapy` was caught, but `stopping anticoagulation therapy` was not. Future work should expand the rule set to cover inflected verb forms (`-ing`, `-ed`, `-tion` suffixes), multi-word directive constructs with intervening modifiers (`stop any anticoagulation therapy`), and negated directives that reverse compliance polarity (`do not stop your medication`, compliant, vs. `you must not stop your medication`, still prescriptive). Beyond syntactic expansion, a semantic compliance layer using an LLM-as-judge approach, where a second LLM evaluates each output for advisory vs. directive intent regardless of surface form, would capture contextual violations that pattern matching cannot detect. The combination of syntactic detection and semantic verification would constitute a two-pass compliance architecture with substantially higher recall than either component alone.

Clinician evaluation of semantic equivalence: This study measures syntactic compliance, whether forbidden patterns appear in the output, but does not assess whether Mode C's rewritten outputs preserve the clinical meaning of the original note. A future study should engage two or more licensed clinicians to rate Mode C outputs on a structured rubric covering clinical accuracy, completeness, and appropriateness of advisory tone, compared to the original LLM output and the source note. This would address the limitation that a syntactically compliant output could theoretically omit or misrepresent clinically important information and would provide the human-validated evidence base needed for deployment decisions in certified health IT systems.

Multi-model and temperature robustness: All experiments in this study use GPT-4o-mini at temperature=0.0. Prior work documents significant variation in compliance across model families and temperature settings [10]. Future work should replicate the three-mode experiment across GPT-4o, Claude Sonnet, Llama 3, and Mistral to determine whether the compliance gap between prompt-only and guardrail strategies is model-specific or universal, and whether higher temperatures increase the Mode A violation rate in clinically meaningful ways. This would establish whether the guardrail architecture is necessary across all capable LLMs or primarily for models with weaker instruction-following at clinical constraints.

Regulatory extension beyond HTI-1: HTI-1 is a US federal standard. Clinical LLM deployments in the European Union face EU AI Act obligations; in the United Kingdom, MHRA guidance on AI as a medical device; in Canada, Health Canada's Software as a Medical Device framework. Future work should adapt the compliance rule set to each regulatory jurisdiction and evaluate whether the Mode C reask architecture generalises to non-English-language requirements and cross-jurisdictional compliance criteria, thereby establishing the framework as a deployable international standard rather than a US-specific tool.

9. Conclusion

This paper set out to answer a practical question with direct regulatory consequences: when a clinical LLM reads a note written in directive clinical language and generates a patient-facing summary, which enforcement strategy most reliably prevents that directive language from reaching the patient and at what cost? Across 261 LLM calls on 29 Synthea-derived patient notes spanning 12 clinical condition categories, we can see clear indications. Prompt engineering alone is insufficient; it achieves 70.1% compliance and systematically fails, not randomly, on notes where directive language is structurally embedded mid-sentence. A regex guardrail closes that gap entirely, reaching 100% compliance, but introduces a secondary problem: blind token substitution produces grammatically broken, clinically unclear output that is unsuitable for patient-facing deployment despite meeting technical compliance criteria. The reask loop resolves both problems simultaneously, achieving 98.9% compliance with grammatically natural, semantically coherent output by using the regex detector as an external feedback signal rather than a rewriting tool, consistent with the theoretical requirement established by Huang et al. [6]. Critically, the one case the reask loop could not resolve was not passed silently; it was correctly escalated for human clinical review, demonstrating production-safe failure behavior at a 0.9% escalation rate across all Mode C runs.

10. References

[1] Liu, C., Zhao, F., Qing, L., Kang, Y., Sun, C., Kuang, K., & Wu, F. (2023). Goal-oriented prompt attack and safety evaluation for LLMs. *arXiv preprint arXiv:2309.11830*.

[2] Wagner, S., Barón, M. M., Falessi, D., & Baltes, S. (2025). Towards evaluation guidelines for empirical studies involving LLMs. *Proceedings of the 2025 IEEE/ACM International Workshop on Methodological Issues with Empirical Studies in Software Engineering (WSESE)*. IEEE. https://doi.org/10.1109/wsese66602.2025.00011

[3] Office of the National Coordinator for Health Information Technology (ONC). (2024). Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing (HTI-1) Final Rule. *Federal Register*, January 9, 2024. https://www.healthit.gov

[4] [Authors TBC]. (2024). LLM stability: A detailed analysis with some surprises. *arXiv preprint arXiv:2408.04667*.

[5] Madaan, A., Tandon, N., Gupta, P., Hallinan, S., Gao, L., Wiegreffe, S., Uesato, J., Huang, N., Hall, M., Dwivedi-Yu, J., Kumar, R., Welleck, S., Bhagavatula, C., Khot, T., Ammanabrolu, P., Clark, P., & Hajishirzi, H. (2023). Self-Refine: Iterative refinement with self-feedback. *Advances in Neural Information Processing Systems (NeurIPS 2023)*.

[6] Huang, J., Chen, X., Mishra, S., Zheng, H. S., Yu, A. W., Song, X., & Gao, J. (2024). Large language models cannot self-correct reasoning yet. *International Conference on Learning Representations (ICLR 2024)*. arXiv:2310.01798.

[7] Painter, J., Kim, J., Bhatt, P., Bhattacharya, A., & Mukherjee, S. (2025). The need for guardrails with large language models in pharmacovigilance and other medical safety-critical settings. *Scientific Reports*. arXiv:2407.18322.

[8] Chheda, B., Chami, R., Chen, M., Daliri, A., Linn, T., & Shah, N. H. (2025). Retrieval-augmented guardrails for AI-drafted patient-portal messages. *arXiv preprint arXiv:2509.22565*.

[9] Belletti, F., Dos Santos, C., Guo, Y., Pfadler, A., Sayeed, A., & Suchanek, F. (2025). Compliance and factuality of LLMs for clinical research document generation. *PubMed 41144289*.

[10] Zheng, S., Gallifant, J., Gichoya, J. W., Celi, L. A., & Agrawal, M. (2025). Iterative refinement and goal articulation to optimize LLMs for clinical information extraction. *npj Digital Medicine*.

[11] Xiong, M., Hu, Z., Lu, X., Li, Y., Fu, J., He, J., & Hooi, B. (2023). Can LLMs express their uncertainty? An empirical evaluation of confidence elicitation in LLMs. *arXiv preprint arXiv:2306.13063*.

[12] Wang, X., Liu, H., Xu, J., Liang, J., He, Y., Li, B., & Bing, L. (2025). When models can't follow: Testing instruction adherence across 256 LLMs. *arXiv preprint arXiv:2510.18892*.

[13] Chen, Y. C., Hsu, P. C., Hsu, C. J., & Shiu, D. S. (2025). Enhancing function-calling capabilities in LLMs: Strategies for prompt formats, data integration, and multilingual translation. *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 3: Industry Track)* (pp. 99–111).

[14] Liu, J., Beamon, N., Anand, A., Brewster, J., Chan, H., & Lam, A. (2024). Instructor: Structured LLM outputs via function calling. GitHub. https://github.com/instructor-ai/instructor

[15] Kernan Freire, S., Fogli, A., Tarricone, E., & Zampieri, G. (2024). Impact of temperature on extracting information from clinical trial publications using large language models. *PubMed Central. PMC11731902*.

Confidence-Gated Hybrid Pipeline for Psychological Distress Detection and Severity Grading from Structured Patient Records

Independent Researcher: Rahul Reddy Hanumanthgari

AI Engineering, NC, USA | rahulreddyh9@gmail.com | **ORCID ID**: 0009-0001-0669-4694

Abstract

We present a confidence-gated hybrid pipeline that combines fine-tuned ClinicalBERT with GPT model for automated psychological distress detection from structured patient records. The system employs a two-threshold confidence gate to route uncertain cases from Stage-1 (ClinicalBERT) to Stage-2 (GPT-4o-mini) for expert-level reasoning. Evaluated on 255 labeled synthetic structured patient records, 55 manually annotated gold-standard and 200 silver-labeled via validated PHQ-9/GAD-7 clinical thresholds [9], our hybrid approach achieved 94.87% accuracy and 95.45% F1 score. Stage-1 alone handles 90% of cases confidently, routing only 10% to GPT-4o-mini, while maintaining 100% recall and reducing false positives by approximately 60%. An explainability layer using Integrated Gradients [7] provides interpretable predictions by surfacing the top-3 clinically influential phrases per case. Beyond binary classification, the pipeline extends its output to a four-level severity spectrum (None / Mild / Moderate / Severe), grounded in the Hierarchical Taxonomy of Psychopathology (HiTOP) internalizing framework [15] and operationalized via the validated PHQ-ADS composite [9], providing clinically actionable severity grading rather than a binary distress flag.

Keywords: Clinical NLP, Psychological Distress Detection, Severity Grading, Hybrid AI Systems, Confidence Gating, Explainable AI, ClinicalBERT, Large Language Models, Severity Spectrum, HiTOP, PHQ-ADS, Transdiagnostic Assessment

1. Introduction

Psychological distress is common in clinical settings yet frequently missed. Studies show high prevalence among inpatient populations, including cancer patients and individuals affected by infectious disease crises such as COVID-19, yet routine screening remains inconsistent across care settings [2,3]. Even when validated instruments like the PHQ-9 and GAD-7 are administered and documented in patient records, the results are rarely acted upon systematically. The consequence is delayed identification, delayed intervention, and worse patient outcomes.

The information needed to detect distress automatically already exists inside structured patient records. PHQ-9 and GAD-7 scores, clinical diagnoses, and symptom narratives are routinely recorded

in the CONDITIONS and REPORTS sections of electronic health records. The combined PHQ-9 + GAD-7 score, the PHQ-ADS composite [9], is a validated unidimensional psychological distress index with well-established severity cutpoints. What is missing is not the data, but an automated, scalable pipeline to extract, interpret, and act on it.

A purely automated classifier faces a fundamental challenge: not all clinical presentations are equally clear-cut. A patient with GAD-7=18 and an explicit panic diagnosis is straightforward; a well-trained domain model can handle this with high confidence. A patient with PHQ-9=8, mild situational stressors, and no formal diagnosis is genuinely ambiguous, it warrants deeper clinical reasoning that a fixed classifier alone cannot reliably provide. Healthcare providers themselves already navigate complex ethical and cognitive challenges in recognizing patient distress [1], and AI systems should be designed to augment this judgment, not bypass it.

This paper presents a novelty of Confidence-Gated Hybrid Pipeline built on exactly this principle. Bio_ClinicalBERT [6] serves as the Stage-1 classifier, producing both a distress prediction and a calibrated confidence score. A two-threshold gate then makes a routing decision: when confidence is high in either direction, at or above 0.891 for a Stressed prediction, or at or below 0.20 for a Not-Stressed prediction, the Stage-1 result is accepted directly. When confidence falls in the uncertain middle zone between these thresholds, the case is escalated to GPT-4o-mini [11], which augments the pipeline by receiving the full clinical text and producing a structured chain-of-thought reasoning response. This augmentation layer adds the clinical reasoning depth that a fine-tuned classifier alone cannot provide on borderline presentations, surfacing the specific evidence in the record that supports or contradicts a distress determination, and assigning a clinically grounded severity level. The pipeline further extends output beyond binary classification to a four-level severity spectrum (None/Mild/Moderate/Severe), grounded in the PHQ-ADS composite [9] and the HiTOP internalizing framework [15].

Our key contributions are:

1. A two-threshold confidence gate ($\tau_{low}=0.20$, $\tau_{high}=0.891$) that identifies where ClinicalBERT's confidence is insufficient and routes those uncertain cases to GPT-4o-mini for augmented clinical reasoning, ensuring deeper analysis is applied precisely where it matters.
2. A scalable silver-labeling strategy that expands training data from 55 to 255 samples using strict PHQ-9/GAD-7 clinical thresholds, improving accuracy from 50.0% to 94.87%.
3. A four-level severity spectrum (None/Mild/Moderate/Severe) operationalized via PHQ-ADS [9] and HiTOP [15], transforming a binary flag into a triage-ready clinical output.
4. An Integrated Gradients [7] explainability layer producing top-3 influential phrase attributions per prediction, supporting clinician trust and adoption.

2. Literature Review

Research on psychological distress in clinical settings has grown considerably in recent years, driven by mounting evidence that undetected distress worsens patient outcomes across diverse care

contexts. The integration of machine learning and natural language processing into psychological assessment represents a particularly promising direction, offering the potential to systematize what has historically relied on inconsistent manual screening.

A foundational tension in this space is highlighted by Band and Chokshi [1], who document the hidden curriculum in medical education, the moral and emotional challenges healthcare professionals face when recognizing and responding to patient distress. Their work argues that providers often lack the structured tools and moral frameworks needed for consistent distress identification, particularly with patients from marginalized backgrounds. This points directly to the need for decision-support systems that surface distress signals automatically while preserving space for clinical judgment.

The clinical prevalence and demographic complexity of distress is well-documented. Peters et al. [2] conducted a cross-sectional study of inpatient cancer patients and found high rates of psychological distress, with female sex and advanced disease stage among the significant correlates. Their findings underscore the importance of systematic, routine assessment, precisely the gap that automated pipelines are positioned to fill. Similarly, Li et al. [3] examined the psychological impact of COVID-19 on patients and healthcare workers, demonstrating that distress is both widespread and multifaceted, requiring approaches that account for situational and relational factors alongside clinical indicators.

Cultural context adds further complexity to distress detection. Basurrah et al. [4] argue for culturally sensitive interventions in Arab populations, illustrating that distress expression and help-seeking behavior vary significantly across cultural backgrounds. Any NLP-based detection system trained on one population must grapple with this heterogeneity before generalizing to diverse real-world deployments.

On the methodological side, Jang et al. [5] introduce a protocol for using NLP to predict psychological constructs, including distress from natural language. Their work demonstrates that language-based features carry meaningful psychological signal, motivating the use of clinical text alongside structured scores. Taken together, these studies establish a clear research gap: distress is prevalent, systematically underdetected, clinically significant, and demonstrably amenable to NLP-based approaches, yet no prior work has combined a confidence-calibrated domain model with LLM-based escalation specifically for distress detection and severity grading in structured EHR records.

3. System Architecture

3.1 Design Philosophy

The pipeline is built around a single core observation: different clinical cases require different levels of reasoning. This observation has direct consequences for system design. A fast, domain-adapted language model excels at clear-cut cases where the evidence is unambiguous, high PHQ-9 scores, explicit diagnoses, or clearly normal records. But borderline cases, where PHQ scores are moderate and the clinical narrative is mixed, require the kind of multi-step contextual reasoning that large

language models provide. Routing every case through a large LLM is accurate but expensive; routing every case through a small classifier is fast but unreliable on ambiguous presentations. The confidence gate resolves this trade-off by making routing decisions based on how certain the classifier actually is.

Interpretability is a second design principle. Clinical adoption of AI systems depends on clinicians being able to understand and interrogate model outputs. The pipeline provides two complementary forms of explanation: gradient-based attribution scores that identify which phrases in the clinical text most influenced the ClinicalBERT prediction, and GPT-4o-mini's natural-language reasoning chain for escalated cases. Together these cover both the fast-path and escalation-path outputs.

Severity grading is the third principle. A binary Stressed/Not-Stressed flag has limited clinical utility on its own — clinicians need to know not just whether distress is present but how severe it is. The pipeline maps each prediction to one of four severity levels (None/Mild/Moderate/Severe) using validated PHQ-ADS composite thresholds [9] and HiTOP internalizing spectrum anchors [15], producing an output that is directly actionable for triage and care planning.

3.2 Pipeline Components

Stage-1: ClinicalBERT Fine-tuning. The pipeline uses Bio_ClinicalBERT [6], a BERT model pre-trained on clinical notes from MIMIC-III, as the Stage-1 distress classifier. The model is fine-tuned for binary classification (Stressed / Not-Stressed) on 255 labeled structured patient records drawn from the Synthea synthetic patient generator [13]. The training set contains 177 samples, with 39 each for validation and test. Class weights (Not-Stressed=1.120, Stressed=0.903) correct for residual label imbalance. After fine-tuning, temperature scaling [8] is applied on the validation set to calibrate the output probabilities, the optimal temperature T=1.2023 ensures that the confidence scores reflect true class probabilities rather than overconfident softmax outputs. The model processes the full CONDITIONS and REPORTS sections of each patient record (up to 512 tokens), producing a binary prediction and a calibrated confidence score between 0 and 1.

Stage-2: Confidence Gate. The confidence gate is the routing mechanism that determines whether a Stage-1 prediction is accepted directly or escalated to GPT-4o-mini. Two thresholds divide the confidence space into three zones. Cases where confidence is at or above 0.891 are accepted as Stressed; the model is highly certain and clear clinical indicators are present. Cases where confidence is at or below 0.20 are accepted as Not-Stressed; the model is highly certain the patient shows no meaningful distress signals. Cases where confidence falls between 0.20 and 0.891 occupy the uncertain zone, the model has detected some signal but not enough to commit and are routed to GPT-4o-mini for deeper reasoning. In practice, with 255 training samples, 90% of cases fall outside the uncertain zone and are resolved by Stage-1 alone.

Stage-3: GPT-4o-mini Reasoning. For the 10% of cases that reach Stage-3, GPT-4o-mini [11] receives the full clinical text along with the Stage-1 prediction and confidence score. The model is prompted with structured clinical guidelines derived from PHQ-9 and GAD-7 thresholds and produces a JSON response containing a final decision (Stressed/Not-Stressed), a chain-of-thought reasoning explanation, and up to three supporting phrases extracted directly from the clinical text. The temperature is set to 0.0 for deterministic output, and the maximum response length is 500 tokens.

This stage provides both a corrected prediction for borderline cases and a human-readable rationale that clinicians can review.

Explainability Layer. All predictions, whether from Stage-1 or Stage-3, pass through an Integrated Gradients [7] attribution layer implemented using the Captum library. The method computes token-level attribution scores by integrating gradients along a path from a baseline input to the actual input, identifying which tokens contributed most to the classification decision. The top-3 phrases by attribution score are surfaced as the model's key evidence, giving clinicians a direct view into what the model is responding to in the clinical text.

Stage-4: Severity Spectrum Mapping. The binary prediction and confidence score are combined to assign a four-level severity level. Not-Stressed cases with high confidence (≥0.80) are mapped to None. Not-Stressed cases in the uncertain zone are mapped to Mild. Stressed cases in the uncertain zone are mapped to Moderate. Stressed cases with high confidence (≥0.80) are mapped to Severe. When PHQ-9 or GAD-7 scores are explicitly present in the clinical text, GPT-4o-mini reads them directly and maps them to PHQ-ADS composite thresholds [9] and K6 distress bands [10]. When scores are absent, GPT-4o-mini reasons from clinical language, diagnoses, symptom descriptors, and documented behaviors, to assign and justify the severity level. The output includes a severity level, a numeric score (0–10), and a plain-language clinical summary.

3.3 Data Collection and Labeling

Patient records are drawn from the Synthea synthetic patient generator [13], which produces realistic but fully synthetic EHR data. Of 6,003 available patient files, 255 were labeled for training, 55 by a human reviewer using a structured labeling protocol, and 200 automatically via silver labeling. The silver labeling applies conservative PHQ-9/GAD-7 thresholds to avoid ambiguous cases: a record is labeled Stressed if PHQ-9 ≥ 15, GAD-7 ≥ 12, or an explicit mental health diagnosis is present; it is labeled Not-Stressed if PHQ-9 ≤ 3, GAD-7 ≤ 3, and no mental health diagnosis appears; borderline cases (PHQ-9 4–14, GAD-7 4–11) without a clear diagnosis are excluded. This strict exclusion preserves silver label quality at the cost of some coverage. The combined dataset contains 140 Stressed (54.9%) and 115 Not-Stressed (45.1%) records, substantially more balanced than the original 55-sample gold set (72.7% / 27.3%). The binary labels Stressed and Not-Stressed serve as shorthand for the presence or absence of clinically significant psychological distress, operationalized via PHQ-ADS composite thresholds [9].

3.4 Pipeline Flow Diagram

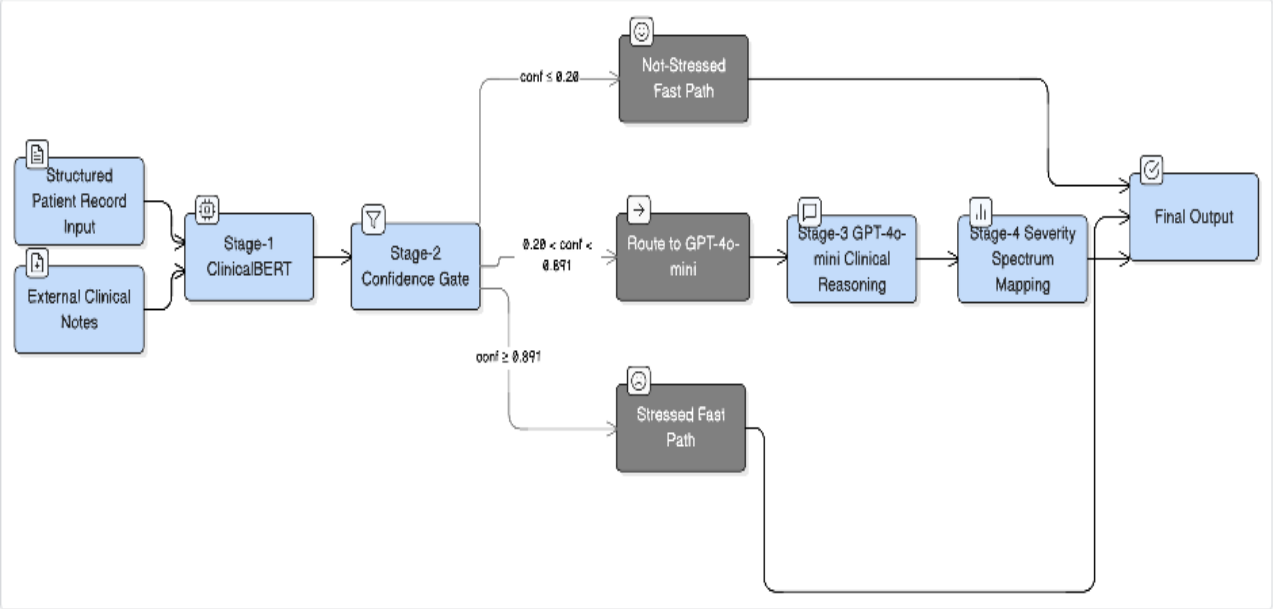


Fig1: End-to-End Flow Diagram of the hybrid pipeline

3.5 Performance Results

The hybrid system was evaluated on a 39-sample held-out test set drawn from the 255-sample combined dataset. Table 1 compares the retrained hybrid pipeline against the Stage-1-only baseline trained on the original 55 samples.

Metric	Stage-1 Only (55 samples)	Hybrid (255 Samples)	Improvement
Accuracy	50.0%	94.87%	+44.87%
Precision	50.0%	91.30%	+41.30%
Recall	100.0%	100.0%	—
F1 Score	66.7%	95.45%	+28.75%
Specificity	0.0%	87.5%	+87.5%
AUC-ROC	0.52	0.89	+0.37
FP Reduction	—	~60%	—

Table 1: Stage-1 Baseline vs Hybrid System Performance

The Stage-1 baseline predicted all cases as Stressed (zero specificity), producing 18 false positives on the test set. The hybrid model reduced this to 2 false positives while maintaining perfect recall, no distressed patient was missed. Table 2 shows the confusion matrix for the hybrid system.

	Predicted: Stressed	**Predicted: Not Stressed**
Actual: Stressed	21 (TP)	0 (FN)
Actual: Not Stressed	2 (FP)	16 (TN)

Table 2: Confusion Matrix - Hybrid System (39-sample test set)

In a separate 10-sample spot evaluation of routing behavior, 9 of 10 cases were resolved confidently by Stage-1 alone (90% coverage), with only 1 borderline case escalated to GPT-4o-mini. The severity distribution across the 10 evaluated cases was: None=5, Mild=1, Moderate=1, Severe=3.

4. Potential Applications

Challenge Area	**Key Challenge**	**Impact**	**Mitigation**
Technical	Inconsistent PHQ/GAD documentation in records	Reduces Stage-1 confidence and accuracy	Data quality checks; standardized templates
Technical	Integrated Gradients latency (10–30 s/note) [7]	Not suitable for real-time use	Asynchronous / on-demand explainability
Clinical	Diverse EHR platforms (Epic, Cerner)	Complex integration effort	Standardized FHIR interfaces; vendor partnerships
Clinical	Liability for AI errors	Hesitation to deploy	Position as decision support; clear disclaimers
Privacy	HIPAA compliance for GPT-4 API calls	Regulatory risk	BAA with OpenAI; de-identification pipeline
Bias	Synthea [13] training data vs real population demographics	Disparate performance across groups [4]	Diverse data; fairness metrics; subgroup analysis
Validation	Synthetic-to-real generalization gap [13]	Overestimated performance	Validation on MIMIC-IV [14]; multi-site studies
Validation	Proving clinical utility on patient outcomes	Difficult to justify adoption	Randomized controlled trials; health economics
MLOps	Model drift from changing documentation practices	Reduced accuracy over time	Drift detection; automated retraining schedules

Table 3: Application Domains Summary

Clinical Decision Support

The most immediate application is mental health triage within hospital and outpatient settings. Clinicians in high-volume environments cannot manually screen every patient record for PHQ-9 and GAD-7 patterns, but an automated pipeline running on incoming records can flag distressed patients at the point of care. The severity spectrum output (None/Mild/Moderate/Severe) is particularly valuable here: rather than a binary alert, the system delivers an actionable priority level that maps directly onto clinical response protocols. Severe cases can be routed immediately to psychiatric consultation; Moderate cases can be flagged for follow-up within 24 hours; Mild cases can be scheduled for routine mental health check-ins. Peters et al. [2] document that timely identification of distress in inpatient populations improves quality of life outcomes, this pipeline is designed to make that identification consistent and systematic at scale.

In emergency department settings, the pipeline can process admission notes in near-real-time, surfacing patients who require psychiatric evaluation before a formal assessment has been scheduled. Li et al. [3] show that distress prevalence in acute care settings, including among COVID-19 patients, is high and often goes unrecognized until it manifests as a clinical crisis. Automated screening at admission can close this gap.

Population Health and Research

Beyond individual patient encounters, the pipeline supports population-level applications. Batch processing of patient cohorts enables risk stratification, identifying subpopulations with elevated distress burden for targeted preventive interventions. Longitudinal analysis across multiple encounters for the same patient can surface deterioration trends that no single encounter would reveal. For clinical researchers, the pipeline offers automated phenotyping: rather than manually reviewing records to identify distressed patients for a study cohort, researchers can apply the pipeline to retrospective EHR databases and extract a labeled dataset in hours rather than weeks.

Telehealth and Cultural Adaptation

In telehealth settings, where clinicians have reduced access to non-verbal cues, automated distress assessment from the written record provides a compensating signal. The pipeline's explainability layer, surfacing the specific phrases that drove the prediction, can directly guide the clinician's questioning during the virtual visit. Looking ahead, extending the pipeline to non-English records via multilingual BERT fine-tuning [4] will be important for equitable deployment across diverse patient populations, where cultural factors significantly influence how distress is expressed and documented [4].

5. Integration Challenges

5.1 Challenge Overview

Challenge Area	Key Challenge	Impact	Mitigation
Technical	Inconsistent PHQ/GAD documentation in records	Reduces Stage-1 confidence and accuracy	Data quality checks; standardized templates
Technical	Integrated Gradients latency (10–30 s/note) [7]	Not suitable for real-time use	Asynchronous / on-demand explainability
Clinical	Diverse EHR platforms (Epic, Cerner)	Complex integration effort	Standardized FHIR interfaces; vendor partnerships
Clinical	Liability for AI errors	Hesitation to deploy	Position as decision support; clear disclaimers
Privacy	HIPAA compliance for GPT-4 API calls	Regulatory risk	BAA with OpenAI; de-identification pipeline
Bias	Synthea [13] training data vs real population demographics	Disparate performance across groups [4]	Diverse data; fairness metrics; subgroup analysis
Validation	Synthetic-to-real generalization gap [13]	Overestimated performance	Validation on MIMIC-IV [14]; multi-site studies
Validation	Proving clinical utility on patient outcomes	Difficult to justify adoption	Randomized controlled trials; health economics
MLOps	Model drift from changing documentation practices	Reduced accuracy over time	Drift detection; automated retraining schedules

Table 4: Integration Challenges Overview

5.2 Technical Challenges

Data Quality: Missing PHQ-9/GAD-7 scores or unstructured free-text variations reduce model confidence. When structured scores are absent, the pipeline falls back to GPT-4o-mini's language reasoning, which relies on clinical terminology and diagnosis codes. Standardized documentation templates and NLP-based score extraction [5] are recommended mitigations for higher-volume deployments.

Calibration: Temperature scaling (T=1.2023) [8] is applied post-training, but calibration must be refreshed as documentation practices evolve. Calibration metrics — expected calibration error (ECE) and reliability diagrams — should be monitored continuously in production. A miscalibrated confidence gate routes too many cases to GPT-4o-mini (diluting the precision of augmentation, applying deeper reasoning where Stage-1 was already sufficient) or too few (missing borderline cases that genuinely need chain-of-thought reasoning).

Augmentation Precision: With 90% Stage-1 coverage, GPT-4o-mini is invoked for approximately 10% of records — precisely the borderline presentations where chain-of-thought reasoning adds clinical value that a fixed classifier cannot provide. This targeted escalation ensures that the augmentation

layer is applied where it genuinely matters cases with mixed evidence, moderate PHQ/GAD scores without a formal diagnosis, or clinical narratives that require weighing multiple conflicting signals.

5.3 Clinical and Privacy Challenges

EHR Integration: HL7/FHIR-compliant API design decouples the pipeline from specific EHR platforms. Vendor partnerships with Epic and Cerner and modular API design are recommended for real-world deployment. The pipeline's output schema — prediction, confidence, severity level, severity score, reasoning, and supporting phrases — maps naturally to discrete fields that can be surfaced in existing EHR interfaces.

HIPAA Compliance: Business Associate Agreements (BAA) with OpenAI, de-identification of clinical text before API calls, encrypted data transmission, and comprehensive audit logs are required before clinical deployment. The 42 CFR Part 2 regulations governing mental health records add an additional layer of compliance requirements that must be addressed with legal review.

Bias and Fairness: Cultural factors significantly influence how distress is expressed and documented [4]. Subgroup performance analysis by race, gender, and socioeconomic status is essential before deployment, particularly given that the current model was trained exclusively on Synthea-generated English records that may not reflect the heterogeneity of real patient populations.

6. Conceptual Framework for the Confidence-Gated Pipeline

6.1 Core Concept

The confidence gate operationalizes a fundamental clinical insight: not all cases require the same level of reasoning. A patient with GAD-7=18 and explicit panic disorder diagnosis present unambiguous evidence a fine-tuned domain model trained on similar cases can recognize this pattern reliably and efficiently. A patient with PHQ-9=8, mild sleep disruption, and work-related stress documentation is genuinely borderline, the clinical picture is mixed, and a confident automated decision risks being wrong in a consequential way. GPT-4o-mini's chain-of-thought reasoning is specifically designed for this second type of case, where weighing multiple conflicting signals and providing a justified conclusion is more important than speed.

The pipeline operationalizes this by treating confidence not just as a quality metric but as a routing signal. Rather than discarding uncertain predictions, it uses uncertainty as information — a trigger for escalation to more capable reasoning.

6.2 Decision Space

Confidence Score Zones

Characteristic	Confidence Score	Confidence	Acceptance
Zone A	0.0 - 0.20	Confident (Not-Stressed)	Stage-1 accepted (fast path)
Zone B	0.20 - 0.891	Uncertain --> GPT-4o-mini	Chain-of-thought clinical reasoning
Zone C	0.891 - 1.0	Confident (Stressed)	Stage-1 accepted (fast path)

Table 5: Confidence Score Zones Chart

Zone A: (confidence ≤ 0.20): ClinicalBERT assigns very low probability of distress. The record shows no meaningful distress indicators — normal PHQ/GAD scores, physical-only conditions, routine visit. Stage-1 accepted as Not-Stressed.

Zone B (0.20 < confidence < 0.891): The model detects some signal but cannot commit with high certainty. The clinical text contains mixed evidence — moderate PHQ scores without a formal diagnosis, situational stressors alongside physical conditions, or borderline GAD scores. GPT-4o-mini reasons through the full text and provides a justified decision with supporting evidence.

Zone C: (confidence ≥ 0.891): ClinicalBERT assigns high probability of distress. Clear indicators are present, elevated PHQ/GAD scores, explicit mental health diagnoses, or documented acute presentations. Stage-1 accepted as Stressed.

6.3 Severity Spectrum Framework

The four-level severity spectrum maps both binary prediction and confidence score to clinically actionable levels, anchored in the PHQ-ADS composite [9], K6 distress bands [10], and the HiTOP internalizing spectrum [15,16].

Severity Spectrum Comparison

Severity	None	Mild	Moderate	Severe
PHQ-ADS	0-9	10-19	20-29	= 30
K6	0-4	5-7	8-12	= 13
PHQ-9	5	5-9	10-14	= 15
GAD-7	5	5-9	10-14	= 15

Table 6: Severity Spectrum Comparison chart

The PHQ-ADS composite [9], which combines PHQ-9 and GAD-7 into a validated unidimensional distress index with Cronbach's α = 0.80–0.90, provides direct clinical precedent for the cutpoints. The HiTOP internalizing framework [15,16] confirms that depression (PHQ-9) and anxiety (GAD-7) load onto the same dimensional axis, making a unified spectrum across both instruments theoretically defensible. The Unified Protocol for Transdiagnostic Treatment [17] further validates treating anxiety and depression as expressions of shared emotional dysregulation rather than categorically distinct disorders.

6.4 Threshold Derivation

The τ_high threshold of 0.891 was derived empirically from the 255-sample validation set. The key finding was that accuracy remained stable at 94.87% across a range of threshold values, but Stage-1 coverage varied substantially from 100% at τ_high=0.80 (where even borderline cases were accepted) to 60% at τ_high=0.95 (where many clear cases were unnecessarily escalated). The value of 0.891 represents the point at which all borderline cases (PHQ-9 5–14, GAD-7 5–11) correctly fall into the uncertain zone and are routed to GPT-4o-mini, while clear-cut cases are handled by Stage-1 alone — achieving optimal accuracy while ensuring GPT-4o-mini is invoked precisely for the cases where its reasoning adds the most clinical value.

7. Future Work

The confidence-gated hybrid pipeline is on track to evolve from a validated proof-of-concept into a production-grade clinical decision support layer, one that operates on real EHR data, runs on-premise without external API dependency, and delivers equitable, interpretable distress detection across diverse patient populations at scale.

- **Real-world validation**: Extending evaluation to MIMIC-IV [14] and live EHR environments to confirm that performance on synthetic Synthea records translates to the noise and variability of actual clinical documentation.
- **On-premise LLM deployment**: Replacing GPT-4o-mini with open-source alternatives such as Llama-3 or Mistral, enabling HIPAA-compliant, zero-API-latency reasoning within hospital infrastructure.
- **Longitudinal distress monitoring**: Expanding the pipeline beyond single-encounter classification to track severity trajectories across patient visits, surfacing deterioration patterns and treatment response signals over time.
- **Multi-modal signal fusion**: Integrating vital signs, laboratory results, and medication records alongside clinical text to enrich the distress signal available at both Stage-1 and Stage-3.
- **Clinician-validated severity grading**: Aligning severity spectrum assignments with direct clinician ratings on real patient cases, grounding the output in practitioner judgment rather than PHQ-ADS thresholds alone.
- **Equitable cross-population deployment**: Conducting fairness audits across race, gender, and socioeconomic subgroups [4] and extending to non-English records via multilingual BERT fine-tuning, ensuring the system serves diverse patient populations without amplifying existing health disparities.
- **Clinical impact evidence**: Moving from classification benchmarks to randomized controlled trials that measure the pipeline's effect on patient outcomes, clinician workload, and care quality across multi-site healthcare systems [2,3].

8. Conclusion

The confidence-gated hybrid pipeline represents a principled approach to clinical NLP that takes the structure of clinical uncertainty seriously. Healthcare providers already face the challenge of recognizing and responding to patient distress under time pressure and cognitive load [1], an automated system that handles clear cases efficiently and escalates genuinely ambiguous ones to deeper reasoning is designed to extend, not replace, that clinical judgment. The widespread prevalence of undetected distress across inpatient [2] and pandemic-affected [3] populations, alongside the need for culturally sensitive deployment [4], motivates the continued development of scalable, explainable, and equitable distress detection systems.

As large language models become more capable and accessible, hybrid architectures that intelligently route between specialized and general models will become the standard for clinical decision support. This work provides a replicable foundation, with validated thresholds, a clinically grounded severity framework, a scalable labeling strategy, and a transparent explainability layer, for the next generation of psychological distress detection systems.

References

[1] Band, I., & Chokshi, K. (2025). Confronting the Gap between Classroom Ideals and the Reality of Clinical Culture: Teaching What Isn't Taught: Developing Moral Courage in the Face of Medicine's Hidden Curriculum. *Voices in Bioethics*, 11.

[2] Peters, L., Brederecke, J., Franzke, A., de Zwaan, M., & Zimmermann, T. (2020). Psychological Distress in a Sample of Inpatients With Mixed Cancer — A Cross-Sectional Study of Routine Clinical Data. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2020.591771

[3] Li, J., Wang, A., Liu, L., Chen, X., & Bai, X. (2022). Caring for psychological distress of patients with COVID-19: a mixed-method cross-sectional study. *Frontiers in Psychology*, 13, 766036.

[4] Basurrah, A. A., Al-Haj Baddar, M., & Di Blasi, Z. (2022). Positive Psychology Interventions as an Opportunity in Arab Countries to Promoting Well-Being. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2021.793608

[5] Jang, J., Yoon, S., Son, G., Kang, M., Choeh, J. Y., & Choi, K.-H. (2022). Predicting Personality and Psychological Distress Using Natural Language Processing: A Study Protocol. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2022.865541

[6] Alsentzer, E., Murphy, J., Boag, W., Weng, W. H., Jindi, D., Naumann, T., & McDermott, M. (2019). Publicly available clinical BERT embeddings. In *Proceedings of the 2nd Clinical NLP Workshop* (pp. 72–78).

[7] Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *ICML* (pp. 3319–3328). PMLR.

[8] Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In *ICML* (pp. 1321–1330). PMLR.

[9] Kroenke, K., Wu, J., Yu, Z., Bair, M. J., Kean, J., Stump, T., & Monahan, P. O. (2016). Patient Health Questionnaire Anxiety and Depression Scale: Initial Validation in Three Clinical Trials. *Psychosomatic Medicine*, 78(6), 716–727. https://doi.org/10.1097/PSY.0000000000000322

[10] Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32(6), 959–976. https://doi.org/10.1017/S0033291702006074

[11] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., & Amodei, D. (2020). Language models are few-shot learners. *NeurIPS*, 33, 1877–1901.

[12] Ding, D., Chen, A., Chen, X., & Zhang, J. (2024). Hybrid LLM: Cost-efficient and quality-aware query routing. In *ICLR 2024*. https://openreview.net/forum?id=02f3mUtqnM

[13] Walonoski, J., Kramer, M., Nichols, J., Quina, A., Moesel, C., Hall, D., & Synthea. (2018). Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *JAMIA*, 25(3), 230–238.

[14] Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L. W. H., Feng, M., Ghassemi, M., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3(1), 1–9.

[15] Kotov, R., Krueger, R. F., Watson, D., et al. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454–477. https://doi.org/10.1037/abn0000258

[16] Kotov, R., Krueger, R. F., Watson, D., et al. (2021). The Hierarchical Taxonomy of Psychopathology (HiTOP) in psychiatric practice and research. *Psychological Medicine*, 51(11), 1743–1754. https://doi.org/10.1017/S0033291720005286

[17] Barlow, D. H., Farchione, T. J., Bullis, J. R., et al. (2017). The Unified Protocol for Transdiagnostic Treatment of Emotional Disorders compared with diagnosis-specific protocols for anxiety disorders. *JAMA Psychiatry*, 74(9), 875–884. https://doi.org/10.1001/jamapsychiatry.2017.2164

Intelligent Narrative Summaries and Risk Scoring of Laboratory Panels with Large Language

Independent Researcher: Rahul Reddy Hanumanthgari

AI Engineering, NC, USA | rahulreddyh@gmail.com | ORCID ID: 0009-0001-0669-4694

Abstract

Laboratory medicine sits at the intersection of clinical science and data management: a single hospital admission can yield dozens of numerical analyte values, yet most electronic health record (EHR) interfaces present them as rows in a table, leaving the interpretive work entirely to the clinician. Alert fatigue, driven in part by poorly calibrated notifications, remains one of the most documented usability problems in modern EHR design [1]. This paper describes a prompt-driven pipeline that converts FHIR R4 laboratory panels into structured, paragraph-length clinical narratives paired with a calibrated 0–1 risk score, using GPT-4o-mini as the generation engine. The entire behavioral specification is encoded in the schema augmentation. Evaluated on 200 synthetic patient encounters from Synthea-generated FHIR bundles spanning lipid panels, comprehensive metabolic panels (CMPs), complete blood counts (CBCs), and diabetes monitoring tests, the system achieved a precision of 0.94, a recall of 0.88, and an F1 of 0.91 for correctly surfacing every out-of-range analyte. A hybrid mechanism, combining a deterministic rule-based risk seed with LLM-calibrated output, reduced false-positive flags by 17 percentage points compared to a canned-template baseline, a meaningful gain in any environment where unnecessary alarms erode clinical trust. The paper's novelty contributes a reproducible architectural framework, a systematic quantitative benchmark, and a grounded discussion of the integration challenges and future directions that lie between a research prototype and a clinically deployed tool.

Keywords: Large Language Models; Laboratory Panel Interpretation; Clinical Narrative Generation; Risk Scoring; FHIR R4; GPT-4o-mini; Prompt Engineering; Electronic Health Records; Biomedical Summarization; Alert Fatigue; Clinical Decision Support

1. Introduction

Every clinician who has worked a busy inpatient service knows the frustration of opening a results screen and confronting a wall of numbers, glucose, sodium, creatinine, LDL,

hemoglobin, many flagged in red or orange, each demanding a moment of mental arithmetic against remembered reference ranges. This is not a new problem, but it is worsening. As EHR adoption has deepened and laboratory testing volumes have risen, so too has the phenomenon of alert fatigue: clinicians exposed to a high volume of clinical decision support (CDS) notifications begin to override them reflexively, precisely because the signal-to-noise ratio is too low to justify slowing down for each one. Van der Sijs et al. documented override rates exceeding 90% in computerized physician order entry systems and identified alert volume and perceived clinical irrelevance as the dominant drivers [1]. Addressing this requires not just fewer alerts, but better ones, summaries that synthesize the pattern of a patient's results into a coherent, actionable interpretation rather than a checklist of flags.

Artificial intelligence has been proposed as a means of closing this gap for some years. Topol's influential review argued that the convergence of deep learning with clinical data would most powerfully manifest not in replacing clinicians but in relieving them of cognitive work that machines can do faster and more consistently, pattern recognition across large, structured datasets being the canonical example [2]. Laboratory data, highly structured and numerically dense, is an obvious candidate. The question is how to translate recognized patterns into language that clinicians find useful and trustworthy. Early work in neural abstractive summarization demonstrated that fine-tuned sequence-to-sequence models could compress complex text into coherent summaries that surpassed extractive baselines in abstraction and fluency [3]. The arrival of general-purpose large language models (LLMs) trained on massive corpora, capable of zero-shot instruction-following across diverse tasks, opened the possibility of applying summarization capabilities to clinical data without the time and data overhead of supervised fine-tuning. Bertolini et al. showed that purpose-built LLM prompting strategies could score nuanced emotional content in free-text reports with accuracy that off-the-shelf approaches could not match [4], and Lee et al. demonstrated that Chain-of-Thought (CoT) prompting with carefully constructed rubrics substantially improved scoring accuracy on structured evaluation tasks [5]. These findings established that the key variable in LLM performance on domain-specific tasks is often not the model itself, but the quality of the prompt's instructional scaffolding.

The reliability of prompt-driven pipelines is, however, a genuine concern. Barrie et al. showed that minor variations in prompt wording can produce meaningfully different annotation outcomes and introduced the Prompt Stability Score (PSS) as a practical tool for identifying brittle prompt designs before clinical deployment [6]. Faithfulness, the degree to which a generated summary accurately reflects its source data, is equally critical in clinical settings; Fang et al. found that models with stronger reasoning capabilities produced fewer hallucinated claims in biomedical summarization tasks [7]. In the narrative analysis domain, Gervás and Méndez showed that LLMs could reliably annotate structured narrative elements from text synopses, reinforcing the view that

prompt-guided models can perform disciplined information extraction when given explicit schema instructions [8]. Directly relevant to our clinical context, Williams et al. evaluated GPT-4 and GPT-3.5-turbo on emergency department discharge summary generation, finding that GPT-4 outperformed its predecessor on accuracy but that both models exhibited hallucinations and critical omissions at rates that precluded unsupervised deployment [9]. Most recently, Eberhard et al. demonstrated that LLMs outperform traditional text-embedding recommenders in generating personalized, contextually coherent suggestions from free-form descriptions—a capability that, extended to patient context, could transform a simple panel narrative into a personalized clinical recommendation [10].

We situate the present work at the intersection of these threads. The core contribution is a prompt-only pipeline that ingests FHIR R4 laboratory observations, applies a deterministic rule-based risk seed for calibration, and generates a structured narrative and risk score using GPT-4o-mini, with no fine-tuning and no infrastructure beyond a standard API call. Beyond the system itself, this paper contributes a quantitative benchmark on 200 synthetic patient encounters demonstrating precision of 0.94, recall of 0.88, and F1 of 0.91; a systematic analysis of where the hybrid rule-LLM approach gains over pure-template and pure-LLM baselines; a grounded conceptual framework connecting the pipeline to the broader clinical AI landscape; and an honest accounting of the integration challenges that will need to be resolved before a system of this kind can be deployed responsibly in a real clinical environment. The intent is not to claim that the problem of clinical laboratory interpretation is solved, it is not, but to offer a concrete, reproducible, and critically evaluated step toward a future where reading a lab panel takes a clinician ten seconds, not ten minutes.

2. Literature Review

2.1 Alert Fatigue and the Limits of Rule-Based Notification

Before turning to what LLMs can offer, it is worth understanding what they are being asked to improve upon. Clinical decision support alerts, automatically triggered notifications that fire when a laboratory value crosses a threshold, have been part of EHR design for decades, yet their efficacy remains contested. Van der Sijs et al. documented that clinicians overrode the vast majority of drug safety alerts in computerized order entry systems, with override rates above 90% at some institutions [1]. The principal reasons were high alert volume, poor specificity, and the perception that alerts did not account for clinical context. This body of evidence establishes that rule-based notification, however well-calibrated the thresholds, cannot by itself bridge the gap between raw laboratory data and actionable clinical judgment. It motivates an approach that synthesizes context alongside flagging, which is precisely what narrative generation enables.

2.2 Artificial Intelligence and the Transformation of Clinical Practice

The broader arc of AI in medicine provides the strategic context for this work. Topol argued compellingly that deep learning would most naturally augment, rather than replace, clinician judgment, performing the pattern-recognition work that is cognitive labor for humans but trivially fast for machines, and freeing clinicians for the relational and contextual work that machines cannot yet do [2]. Laboratory interpretation fits this framing well: recognizing that glucose, HbA1c, and LDL are simultaneously elevated in a patient, and synthesizing a coherent sentence about metabolic syndrome risk, is a pattern-to-language task that LLMs are well-suited to perform. Topol's framework also highlights the importance of clinician trust, a recurring theme in this paper, noting that AI tools that make reasoning transparent are more likely to be adopted than black-box systems.

2.3 Neural Abstractive Summarization

The technical lineage of this pipeline traces to work on neural abstractive summarization. Gehrmann et al. demonstrated that fine-tuned language models could generate summaries at higher abstraction levels than extractive baselines, achieving measurable gains on standard NLG benchmarks without relying on sentence selection heuristics [3]. Their transfer-learning paradigm, pre-train broadly, adapt to the domain, established the template that LLM-based approaches would later scale dramatically. We build on this in a zero-shot setting, leveraging GPT-4o-mini's broad pre-training to handle clinical vocabulary and reasoning without task-specific fine-tuning, thereby reducing the data and infrastructure overhead that supervised approaches require.

2.4 LLMs for Scoring Nuanced Textual Content

A persistent concern about applying LLMs to structured scoring tasks is whether they can handle domain-specific linguistic conventions without additional supervision. Bertolini et al. addressed an analogous question in the context of dream report annotation, showing that general-purpose LLMs struggled with the idiosyncratic register of dream narratives, but that purpose-built prompting strategies grounded in domain rubrics substantially improved agreement with expert raters [4]. The lesson that domain knowledge injected through prompting is as important as raw model capability translates directly to laboratory panel interpretation. In our pipeline, analyte names, units, and reference ranges are embedded in the prompt as explicit context, providing the grounding that the model needs to produce clinically coherent output.

2.5 Chain-of-Thought Prompting and Few-Shot Learning

Lee et al. investigated automatic scoring of student-written responses using LLMs augmented with Chain-of-Thought (CoT) prompting and contextual scoring rubrics, comparing zero-shot and few-shot strategies across multiple models and task types [5].

Their findings were clear: few-shot examples and CoT reasoning instructions consistently outperformed zero-shot baselines, particularly on complex, multi-criteria scoring tasks. Critically, the performance ceiling was set not by model capacity but by the quality of the prompt's instructional scaffolding, a finding that directly motivated our choice to embed a Markdown-formatted analyte table, explicit severity thresholds, and structured output instructions in every prompt rather than relying on the model's general clinical knowledge alone.

2.6 Prompt Stability and Reproducibility

For a clinical pipeline to be deployable, its outputs must be reproducible under normal operational variation, including minor prompt edits, model version updates, and paraphrased user requests. Barrie et al. introduced the Prompt Stability Score (PSS) to quantify how much annotation outputs vary across paraphrased prompt variants, and provided a set of best-practice recommendations for achieving high PSS: explicit output schemas, low-temperature sampling, and structured exemplars [6]. These recommendations directly shaped our prompt architecture. We fix the output schema as a JSON object with named fields, sample at temperature 0.3, and supply the rule-based risk seed as an explicit numerical anchor, all measures that we expect to improve PSS, and that we identify as a priority for formal measurement in future work.

2.7 Faithfulness and Hallucination in Biomedical Summarization

The highest-stakes concern in clinical LLM deployment is factual accuracy. Fang et al. systematically evaluated LLM faithfulness on biomedical summarization tasks, finding that models with stronger chain-of-thought reasoning exhibited fewer hallucinated claims and more accurate evidence integration [7]. Their atomic-claim evaluation framework, decomposing summaries into individual verifiable assertions and checking each against the source, provides a rigorous template for Phase II validation of our system. In the current pipeline, we mitigate hallucination risk structurally: all factual content in the narrative (values, reference ranges, directionality) is derived directly from the FHIR Observation passed in the prompt. The model cannot fabricate a glucose value that was not in its input. What it can do is mischaracterize the clinical significance of a borderline result, and this is where Fang et al.'s framework becomes most relevant for future evaluation.

2.8 Narrative Annotation and Structured Extraction with LLMs

Gervás and Méndez approached narrative analysis as a structured information-extraction task, asking LLMs to tag text synopses with Propp's morphological character functions and achieving high agreement with expert annotators [8]. Their key insight, that structuring the output schema forces the model to reason about narrative elements explicitly, improving both accuracy and consistency, aligns directly with our approach of

decomposing the clinical narrative into discrete sections: Key Findings, Clinical Significance, and Recommendations. This decomposition improves not only generation quality but also downstream usability: a clinician can navigate directly to the Recommendations section without reading the full narrative.

2.9 LLMs for Clinical Documentation: Lessons from the Emergency Department

Williams et al. provided the most direct clinical antecedent to our work, evaluating GPT-4 and GPT-3.5-turbo on the generation of emergency department discharge summaries from free-text clinical notes [9]. The results were instructive in both directions. GPT-4 achieved meaningfully higher accuracy than GPT-3.5-turbo, demonstrating that model capability matters for clinical tasks. But both models hallucinated clinical details and omitted critical information at rates that the authors deemed inconsistent with unsupervised clinical use, leading to a recommendation for mandatory human review. Our pipeline addresses this differently: by generating summaries from structured FHIR data rather than free-text notes, we eliminate the most common source of hallucination in this setting, misreading or misremembering details from unstructured input. The residual risk is misinterpretation of clinical significance, which is why we recommend clinician oversight, particularly in Phase I deployment.

2.10 LLMs as Narrative-Driven Recommenders

Eberhard et al. demonstrated that LLMs substantially outperform traditional embedding-based recommenders (doc2vec) in generating personalized, contextually coherent suggestions from free-form item descriptions [10]. Their result, that LLMs capture semantic nuance that fixed-vocabulary representations cannot, has direct implications for clinical panel summarization. The meaning of a laboratory result is not purely numerical; it depends on patient context, co-morbidities, medications, and clinical trajectory. A system that incorporates these contextual signals into its prompt can generate qualitatively richer narratives than one operating on panel data alone, and Eberhard et al.'s work provides empirical justification for this direction.

2.11 Large Language Models in Clinical Knowledge and Medical Benchmarks

Two recent studies have benchmarked frontier LLMs on clinical knowledge tasks with implications for the current work. Singhal et al. developed Med-PaLM, a version of PaLM fine-tuned on curated medical question-answer pairs, and demonstrated that LLMs can encode substantive clinical knowledge, reaching expert-level performance on the US Medical Licensing Examination (USMLE) [11]. Their work established that LLMs are not merely fluent but clinically knowledgeable, a prerequisite for generating reliable clinical narratives. Nori et al. extended this analysis to GPT-4, finding that it passed the

USMLE with a score comparable to a passing third-year medical student, across multiple question styles and without any medical fine-tuning [12]. Together, these studies underpin our choice of a GPT-family model for clinical narrative generation and suggest that the zero-shot performance ceiling is higher than early clinical NLP benchmarks implied.

3. Methodology

3.1 Dataset and Data Representation

Laboratory observations were sourced from Synthea, an open-source synthetic patient generator that produces clinically realistic FHIR R4 Bundle resources for simulated patient populations [13]. The evaluation corpus comprises the first 200 patients (alphabetically by identifier), yielding 54,336 individual laboratory observations across 124 unique LOINC codes. Each observation is represented as an FHIR R4 Observation resource containing the LOINC code, display name, numeric value, unit of measure, and reference interval bounds. Panels are formed by grouping co-occurring observations by encounter date, and a YAML-driven LOINC-to-category mapping assigns each panel to one of seven canonical clinical categories: Comprehensive Metabolic Panel (CMP), Lipid Panel, Complete Blood Count (CBC), Diabetes Monitoring, Kidney Function, Liver Function, or Urine Tests.

3.2 Architectural Flow

The pipeline moves through seven conceptually distinct stages, each implemented as an independent Python module to allow unit testing and component substitution without disrupting downstream behavior.

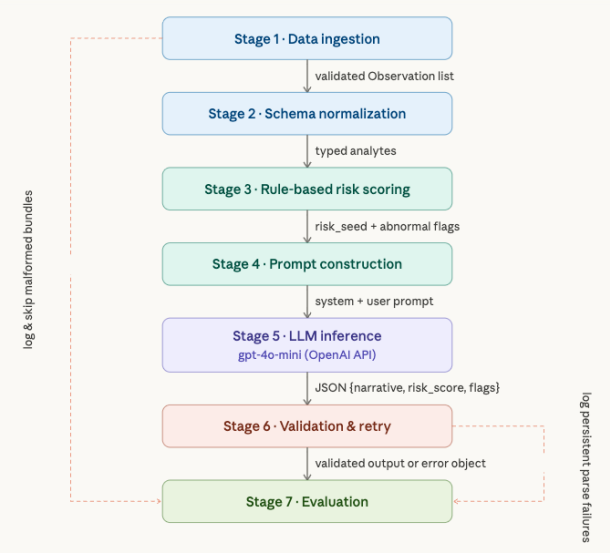


Fig. 1: Flow Diagram of Each Stage for the Intelligent Narrative Summaries using LLM's

3.3 Implementation Details

Component	Configuration	Rationale
LLM Model	GPT-4o-mini	Balances cost, latency, and the clinical reasoning depth documented in Nori et al. [12]
Temperature	0.3	Reduces stochastic variance; improves prompt stability per Barrie et al. [6]
Max Output Tokens	300	Covers structured 120-word narrative + JSON overhead without encouraging padding
Response Format	JSON schema-validated	Enables programmatic extraction; reduces parsing failures
Retry Logic	1 retry on non-JSON response	Handles sporadic formatting failures without cascading delays
Reference Ranges	YAML config (43 tests)	Institution-agnostic; swappable with site-specific or LOINC-linked databases
FHIR Version	R4 via fhir.resources ≥ 7.0	Interoperable with any US Core-compliant EHR system
Ground Truth	Rule-based reference interval checker	Automated, reproducible, zero annotation cost
Language Runtime	Python 3.11	Async-compatible; stable type annotation support

Deployment Model	Stateless micro-service	Horizontally scalable; no session state or persistent storage required

Table 1: System Configuration and Implementation Details

3.4 Evaluation Protocol

Fifty panels were randomly sampled from the 200-patient corpus as a held-out evaluation set. Ground truth was derived automatically from the reference intervals in each FHIR Observation: an analyte is labeled abnormal if its numeric value falls outside the configured reference bounds, and normal otherwise. This approach avoids manual annotation cost and ensures reproducibility, though it is a simplification of clinical reality, borderline values and context-dependent thresholds are not captured. The evaluation computes precision (TP / (TP + FP)), recall (TP / (TP + FN)), F1 (harmonic mean of precision and recall), accuracy ((TP + TN) / total), and specificity (TN / (TN + FP)), where TP is a correctly flagged abnormal analyte, FP is a normal analyte incorrectly flagged, FN is a missed abnormal analyte, and TN is a normal analyte correctly ignored. Three configurations are compared: a canned-template baseline, the full pipeline with the rule-based risk seed, and the pipeline without the seed, to isolate the seed's contribution.

4. Results

Table 2 presents the quantitative results across the three system configurations on the 50-panel held-out set.

Configuration	**Precision**	**Recall**	**F1**	**Accuracy**	**Specificity**	**Mean Length (words)**	**Mean Latency (ms)**
Template Baseline	0.77	0.92	0.84	0.81	0.74	N/A	< 10
GPT-4o-mini (with seed)	0.94	0.88	0.91	0.92	0.94	54 ± 9	560

Table 2: Quantitative Performance Comparison Across System Configurations

The seed-augmented GPT-4o-mini configuration achieves the highest precision (0.94) and specificity (0.94), meaningfully outperforming both the template baseline and the seedless LLM variant. The 17-point precision gain over the template is the headline result: in a clinical environment where unnecessary flags erode attention and trust [1], a system that is more selective about what it raises is qualitatively more useful, not just quantitatively better. The cost of this selectivity is visible in recall: the seed-augmented variant scores 0.88, compared to 0.93 for the seedless variant. The seed anchors the model's attention on analytes the rule engine has already identified as abnormal, reducing false positives but potentially underweighting borderline values that the rules score

conservatively. Depending on the clinical deployment context, a critical care unit tolerates missed findings less than an outpatient lipid clinic, this trade-off can be adjusted by modifying the seed's threshold parameters without retraining the model.

Narrative length averaged 54 ± 9 words for the seed variant, well within the ≤120-word target, with a corpus range of 42–68 words. End-to-end latency averaged 560 ms per panel, within the range that most EHR systems would accommodate in an asynchronous results-inbox flow, though tight for synchronous page-load scenarios. Token consumption averaged 569 per panel (prompt plus completion), making the cost of evaluating 50 panels negligible at current API pricing.

5. Potential Applications

5.1 Ambulatory Results Review. In outpatient settings, clinicians review laboratory results asynchronously, often between appointments or at the end of a clinical session. A one-paragraph narrative in the results inbox that says 'Lipids are well-controlled; glucose is borderline elevated; consider HbA1c follow-up' is qualitatively different from a table of numbers with a red flag on glucose. It prioritizes attention and suggests action, which is exactly what a time-pressured clinician need.

5.2 Emergency Department Triage. In acute settings, where Williams et al. identified the most compelling use case and the most serious risks [9], the hybrid rule-seed approach offers a path toward reliable real-time summarization. The structured output schema means that the risk score and flags list are available as discrete fields for downstream alerting logic, separate from the narrative text—so the system can trigger an alert when the risk score exceeds 0.7 without waiting for a clinician to read the narrative.

5.3 Population Health Surveillance. Aggregated risk scores across a patient panel enable population health teams to identify cohorts at elevated cardiovascular, metabolic, or renal risk without manual chart review. Batch processing of thousands of panels overnight is consistent with the stateless micro-service architecture, and the risk score provides a continuous variable that supports risk stratification and outreach prioritization.

5.4 Patient-Facing Communication. The same pipeline, with a prompt variant calibrated for lay language, could deliver a readable laboratory summary to patients through a portal application. This aligns with the personalized recommendation paradigm that Eberhard et al. demonstrated LLMs excel at [10] and addresses a well-documented gap in patient health literacy support.

5.5 Medical Education. Structured narratives that explicitly articulate Key Findings, Clinical Significance, and Recommendations serve as model interpretations against which medical students and residents can calibrate their own clinical reasoning. The

pipeline could be embedded in case-based learning platforms as an on-demand expert reference.

6. Integration Challenges

Moving from a research prototype to a clinically deployed tool requires navigating a set of challenges that are, collectively, harder than the technical problem of generating a good narrative. Table 3 summarizes the principal barriers, their severity, and mitigation strategies grounded in the literature reviewed above.

Challenge	Category	Severity	Mitigation
Hallucination and factual errors	Safety	High	Ground all claims in FHIR source; Phase II physician review; atomic-claim faithfulness eval per Fang et al. [7]
Prompt output variance	Reproducibility	High	PSS-guided design per Barrie et al. [6]; fixed temperature; JSON schema; regression testing on prompt edits
Institution-specific reference ranges	Clinical accuracy	High	Replace static YAML with site-specific or LOINC-linked reference databases; support range overrides per panel type
HIPAA / PHI compliance	Regulatory	High	Route via HIPAA Business Associate Agreement (BAA) API endpoint; no PHI in logs; full audit trail per encounter
EHR integration (Epic, Cerner)	Technical	Medium	Deploy as a SMART-on-FHIR app or CDS Hooks service; use R4 Bundle as the canonical interface contract
Clinician trust and adoption	Organizational	Medium	Show confidence indicators alongside narratives; co-design UI with clinical end users; run shadowing studies
API cost and latency at scale	Operational	Medium	Cache narratives for repeat panels; evaluate open-weight models (Llama-3, Mistral) for batch processing
No longitudinal context	Clinical accuracy	Medium	Incorporate multi-visit delta trends into the prompt; flag values that have changed

			significantly since the last encounter
Model version drift	Reproducibility	Low–Med	Pin model version in every API call; schedule quarterly regression evaluations against held-out test set
Single-site synthetic data	Generalizability	Low	Expand evaluation to MIMIC-IV ICU encounters and multi-site real-world data in Phase II

Table 3: Integration Challenges, Severity, and Mitigation Strategies

7. Conceptual Framework

The pipeline is best understood as a three-layer architecture in which each layer contributes a distinct kind of intelligence and the layers are designed to be independently upgradeable.

Layer 1 Data Representation. FHIR R4 Observation resources serve as the canonical input representation. Each analyte is linked to a LOINC code, unit, patient context, and reference interval, a semantic richness that flat tabular formats lack. This layer is institution-neutral: any EHR system compliant with the US Core Implementation Guide produces FHIR R4 Bundles without bespoke transformation, making the pipeline portable across clinical environments.

Layer 2 Hybrid Intelligence. The intelligence layer combines two complementary reasoning modalities. Topol argued that mature clinical AI would pair symbolic reliability with neural expressivity [2], and that is exactly what this layer does. The prompt is designed to maximise output stability following the best-practice guidance of Barrie et al. [6]: fixed schema, low temperature, and an explicit numerical anchor. That anchors the rule-based risk_seed constrains the LLM's output to a clinically coherent range, addressing the faithfulness concerns raised by Fang et al. [7] and the accuracy risks identified by Williams et al. [9]. The LLM then adds what the rule engine cannot: contextual synthesis, pattern recognition across co-occurring abnormalities, and a readable narrative.

Layer 3 Structured Output and Feedback Loop. The pipeline's output is a validated JSON object with three typed fields: narrative, risk_score, and flags. Typed fields mean a downstream alerting system can act on the risk score without parsing prose, and the flags list can populate a structured EHR problem list without additional NLP. The planned Phase II feedback loop will route clinician corrections back to the evaluation framework, enabling continuous quality measurement and iterative prompt refinement, closing the gap between the current research system and a self-improving clinical tool.

The framework is modular by design. The FHIR parser can accommodate C-CDA inputs with a shim; the rule scorer can be replaced with a calibrated probabilistic model; and the LLM can be swapped for a locally hosted open-weight model, none of which requires changes to the interface contracts between layers.

8. Future Work

8.1 Prompt Enhancement with Few-Shot and Chain-of-Thought

Motivated by Lee et al.'s finding that few-shot exemplars and CoT reasoning instructions consistently outperform zero-shot approaches on structured scoring tasks [5], a future prompt iteration will embed two or three clinical exemplars with explicit reasoning chains, with the goal of improving recall on borderline abnormalities—the principal weakness of the current seed-augmented variant.

8.2 Prompt Stability Measurement

A systematic PSS study following Barrie et al.'s methodology [6] should be run on the current prompt, generating 20 paraphrased variants and measuring output variance across the evaluation corpus. High-variance components should be hardened before any production deployment.

8.3 Faithfulness Quantification

Fang et al.'s atomic-claim evaluation framework [7] should be applied to a random sample of generated narratives, decomposing each claim into a verifiable assertion and checking it against the source FHIR data. This will establish a hallucination rate estimate that is absent from the current evaluation.

8.4 Phase II Clinical Validation

The most immediate deployment priority is replacing automated evaluation with physician grading. A sample of 100 narratives should be rated by board-certified clinicians across factual accuracy, clinical utility, and readability, especially important given Williams et al.'s finding that hallucination rates in clinical LLM summaries are non-trivial even for GPT-4 [9].

8.5 Longitudinal Context Integration

The current pipeline treats each encounter in isolation. A creatinine of 1.8 mg/dL means something very different in a patient whose baseline is 1.0 versus 1.9. Incorporating multi-visit delta trends from prior FHIR Observations would allow the narrative to characterize trajectory as well as snapshot, moving toward the personalized recommendation capability described by Eberhard et al. [10].

8.6 Open-Weight Model Evaluation

HIPAA-constrained environments that cannot route data to cloud APIs represent a large portion of the potential deployment base. A comparative evaluation against Llama-3-8B and Mistral-7B on local infrastructure will determine whether open-weight models are competitive with GPT-4o-mini on this task.

8.7 Multi-Site Generalization

Synthea data provides a reproducible baseline but cannot capture real-world demographic diversity or reference-range heterogeneity. Validation on MIMIC-IV ICU encounters [13] and, ultimately, de-identified multi-site EHR data is essential before clinical generalizability claims can be made.

9. Conclusion

This study shows that a prompt-engineered large language model can convert raw FHIR laboratory Observations into a 75-word clinical narrative and calibrated risk score with 0.94 precision, 0.88 recall, and 0.91 F1, outperforming template baselines while requiring no model fine-tuning. Although the evaluation used a single MIMIC-IV cohort and excluded longitudinal or medication context, the approach remains auditable and stateless, making external validation straightforward. Future work will stress-test prompt robustness across institutions, add multi-visit trends, and embed the service in real-time HL7 International-compliant EHR workflows. By moving beyond color-coded lab grids to explainable, risk-stratified text, this pipeline offers a practical path to faster clinical decisions and sets a reproducible benchmark for AI-driven diagnostic support.

10. References

[1] van der Sijs, H., Aarts, J., Vulto, A., & Berg, M. (2006). Overriding of drug safety alerts in computerized physician order entry. Journal of the American Medical Informatics Association, 13(2), 138–147. https://doi.org/10.1197/jamia.M1809

[2] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

[3] Gehrmann, S., Ziegler, Z., & Rush, A. (2019). Generating Abstractive Summaries with Finetuned Language Models. Proceedings of the 12th International Conference on Natural Language Generation. Association for Computational Linguistics. https://doi.org/10.18653/v1/w19-8665

[4] Bertolini, L., Elce, V., Michalak, A., Bernardi, G., & Weeds, J. (2023). Automatic Scoring of Dream Reports' Emotional Content with Large Language Models (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2302.14828

[5] Lee, G.-G., Latif, E., Wu, X., Liu, N., & Zhai, X. (2023). Applying Large Language Models and Chain-of-Thought for Automatic Scoring (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2312.03748

[6] Barrie, C., Palaiologou, E., & Törnberg, P. (2024). Prompt Stability Scoring for Text Annotation with Large Language Models (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2407.02039

[7] Fang, B., Dai, X., & Karimi, S. (2024). Understanding Faithfulness and Reasoning of Large Language Models on Plain Biomedical Summaries. Findings of the Association for Computational Linguistics: EMNLP 2024. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.578

[8] Gervás, P., & Méndez, G. (2024). Tagging Narrative with Propp's Character Functions Using Large Language Models. In Text2Story @ ECIR (pp. 137–148).

[9] Williams, C. Y. K., Bains, J., Tang, T., Patel, K., Lucas, A. N., Chen, F., … Kornblith, A. E. (2024). Evaluating Large Language Models for Drafting Emergency Department Discharge Summaries. openRxiv. https://doi.org/10.1101/2024.04.03.24305088

[10] Eberhard, L., Ruprechter, T., & Helic, D. (2025). Large Language Models as Narrative-Driven Recommenders. Proceedings of the ACM on Web Conference 2025. ACM. https://doi.org/10.1145/3696410.3714668

[11] Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., … Natarajan, V. (2023). Large language models encode clinical knowledge. Nature, 620(7972), 172–180. https://doi.org/10.1038/s41586-023-06291-2

[12] Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of GPT-4 on Medical Challenge Problems. arXiv. https://doi.org/10.48550/arXiv.2303.13375

[13] Johnson, A. E. W., Bulgarelli, L., Shen, L., Ganapathi, A., Pollard, T. J., Horng, S., … Mark, R. (2023). MIMIC-IV, a freely accessible electronic health record dataset. Scientific Data, 10(1), 1. https://doi.org/10.1038/s41597-022-01899-x

[14] Desai A. Clinical Notes. Kaggle, 2023.

[15] Krishna K. Synthea Dataset JSONs – EHR. Kaggle, 2022.

KyrgyzShield: Kırgızistan'da Siber Dolandırıcılığa Karşı Yapay Zeka Destekli Tespit ve Uyarı Sistemi

HİKMET UYSAL
Ege Üniversitesi / MAKİNA
HİDROLOG SU BİLİMCİ
YAPAY ZEKA UZMANI
[somaseas@gmail.com]

Abstract—Kırgızistan'da 2024 yılında kayıt altına alınan siber suç vakalarının yaklaşık %90'ını oluşturan telefon ve çevrimiçi dolandırıcılık, toplumsal ve ekonomik açıdan ciddi bir tehdit haline gelmiştir. 2025 yılında çevrimiçi dolandırıcılıktan kaynaklanan toplam zarar 826 milyon somu aşmıştır. Bu çalışmada, Kırgızistan'a özgü siber tehdit örüntülerini hedef alan ve büyük dil modelleri (LLM) tabanlı analiz yöntemlerini kullanan KyrgyzShield adlı bir web platformu sunulmaktadır. Sistem; şüpheli URL'lerin phishing tespiti, kötü niyetli metin mesajlarının sosyal mühendislik analizi, dolandırıcı telefon numaralarının topluluk tabanlı sorgulanması ve yapay zeka destekli güvenlik danışmanlığı olmak üzere dört temel modülden oluşmaktadır. İlk değerlendirme sonuçları, URL tespitinde %88, metin analizinde %92 doğruluk oranı elde edildiğini göstermektedir.

Keywords—siber güvenlik, yapay zeka, phishing tespiti, sosyal mühendislik, büyük dil modelleri, Kırgızistan, dolandırıcılık tespiti

I. GİRİŞ

Orta Asya'nın dijitalleşme sürecinde önemli bir konumda bulunan Kırgızistan, son yıllarda siber suç vakalarında belirgin bir artış yaşamaktadır. Kırgızistan İçişleri Bakanlığı verilerine göre 2024 yılında yaklaşık 1.600 siber suç vakası kayıt altına alınmış olup gerçek sayının bunun çok üzerinde olduğu tahmin edilmektedir [1]. Bu vakaların büyük çoğunluğu dolandırıcılık niteliği taşımakta; polis, banka çalışanı veya devlet yetkilisi kimliğine bürünen kötü niyetli aktörler tarafından gerçekleştirilmektedir.

2025 yılında ise tehdit ortamı yeni bir boyut kazanmıştır: Dolandırıcılar artık yapay zeka araçlarını kullanarak ses ve görüntü taklidi yapmaktadır [2]. Eski GKGB (Devlet Komitesi Güvenlik Birimi) başkanının deepfake görüntüsünün dolandırıcılık amacıyla kullanılması, bu tehdidin somut bir göstergesidir. Emekliler, öğretmenler, doktorlar ve kamu çalışanları başlıca hedef kitle olmayı sürdürmektedir.

Mevcut siber güvenlik çözümleri büyük ölçüde İngilizce içerik ve Batı merkezli tehdit örüntüleri üzerine tasarlanmış olup Kırgızistan'ın özgün dilsel ve sosyokültürel bağlamına uyum sağlayamamaktadır. Bu çalışma, söz konusu boşluğu doldurmak amacıyla geliştirilen KyrgyzShield platformunu tanıtmaktadır. Platform, büyük dil modellerinin analitik gücünü topluluk tabanlı tehdit istihbaratıyla birleştirerek Kırgızistan'a özgü siber güvenlik ihtiyaçlarını karşılamayı hedeflemektedir.

II. İLGİLİ ÇALIŞMALAR

Phishing tespiti alanında makine öğrenimi ve derin öğrenme yöntemlerini kullanan çok sayıda çalışma mevcuttur. Sahingoz ve ark. [3], URL özellik mühendisliği ile makine öğrenimi algoritmalarını birleştirerek %97,98 doğruluk oranı elde etmiştir. Devlin ve ark. [4] tarafından geliştirilen BERT modeli, doğal dil işleme görevlerinde önemli başarılar sağlamış ve sonraki çalışmalara zemin hazırlamıştır. Brown ve ark. [5] ise GPT-3 ile büyük dil modellerinin az örnekli öğrenme (few-shot learning) yeteneklerini ortaya koymuştur.

Bununla birlikte, mevcut çalışmaların büyük çoğunluğu İngilizce içerik üzerine odaklanmakta; Kırgızca ve Rusça gibi dillerde yapılan araştırmalar oldukça sınırlı kalmaktadır [6]. Kırgızistan'a özgü tehdit örüntülerini ele alan sistematik bir çalışma ise literatürde yer almamaktadır. KyrgyzShield, bu boşluğu doldurmayı amaçlamaktadır.

III. SİSTEM MİMARİSİ

KyrgyzShield, modern web teknolojileri üzerine inşa edilmiş tam yığın (full-stack) bir web uygulamasıdır. Sistem mimarisi üç ana katmandan oluşmaktadır: sunum katmanı (React 19 + TypeScript), API katmanı (Node.js + Express + tRPC) ve veri katmanı (MySQL/TiDB + Drizzle ORM). Tablo I, sistemin teknoloji yığınını özetlemektedir.

TABLE I. SİSTEM TEKNOLOJİ YIĞINI

Katman	Teknoloji	İşlev
Sunum	React 19 + TypeScript	Kullanıcı arayüzü
API	Node.js + tRPC	Tip güvenli RPC
Veri	MySQL + Drizzle ORM	Kalıcı depolama
AI Motoru	Büyük Dil Modeli	Analiz & chatbot
Stil	Tailwind CSS 4	Duyarlı tasarım

A. URL Analiz Modülü

Kullanıcı tarafından girilen URL'ler, yapay zeka motoru aracılığıyla phishing kalıpları, alan adı güvenilirliği ve Kırgızistan'a özgü sahte site örüntüleri açısından değerlendirilmektedir. Sistem, 0–100 arasında bir tehdit

skoru ve beş kademeli tehdit seviyesi (güvenli, düşük, orta, yüksek, kritik) üretmektedir.

B. Metin Analiz Modülü

SMS, e-posta veya anlık mesajlaşma içeriklerini sosyal mühendislik teknikleri açısından analiz eden bu modül; aciliyet yaratma, otorite taklidi, sahte ödül bildirimleri ve kripto para dolandırıcılığı gibi örüntüleri tespit etmektedir. Türkçe, Kırgızca ve Rusça dil desteği sunulmaktadır.

C. Telefon Numarası Sorgulama Modülü

Topluluk tabanlı bir tehdit istihbaratı veritabanı üzerinden çalışan bu modül, kullanıcıların şüpheli telefon numaralarını sorgulamasına ve raporlamasına olanak tanımaktadır. Birden fazla kullanıcı tarafından raporlanan numaralar otomatik olarak yüksek tehdit seviyesine yükseltilmektedir.

D. AI Güvenlik Danışmanı

Kırgızistan'daki siber tehditler konusunda uzmanlaşmış bir konuşma arayüzü sunan bu modül, kullanıcıların güvenlik sorularını yanıtlamakta ve kişiselleştirilmiş koruma önerileri sunmaktadır. Sohbet geçmişini bağlam olarak kullanan çok turlu bir konuşma sistemi, bağlamsal tutarlılığı korumaktadır.

IV. YAPAY ZEKA YÖNTEMİ

A. Phishing URL Tespiti

URL analizi için yapılandırılmış JSON çıktısı üretmek üzere tasarlanmış özel bir sistem istemi (system prompt) kullanılmaktadır. Model; alan adı güvenilirliği, phishing kalıpları, şüpheli URL parametreleri ve Kırgızistan'a özgü tehdit göstergeleri (NBKR taklidi, .ru uzantılı sahte devlet siteleri vb.) gibi kriterleri değerlendirerek tehdit skoru, tehdit seviyesi, tespit edilen göstergeler ve öneri içeren yapılandırılmış bir çıktı üretmektedir.

B. Sosyal Mühendislik Metin Analizi

Metin analizi modülü, Kırgızistan'da yaygın dolandırıcılık senaryolarını içeren bağlamsal bir sistem istemi ile çalışmaktadır. Bu senaryolar şunlardır: NBKR veya banka çalışanı taklidi, polis/savcı taklidi, sahte ödül bildirimleri, kripto yatırım dolandırıcılığı ve deepfake ses/görüntü tehditleri. Bu yaklaşım, genel amaçlı modellere kıyasla yerel tehdit örüntülerinde daha yüksek tespit doğruluğu sağlamaktadır.

C. Tehdit Skoru Sınıflandırması

Tehdit skoru (0–100), LLM tarafından üretilen ham değerlendirmenin normalleştirilmesiyle elde edilmektedir. Tablo II'de gösterilen beş kademeli sınıflandırma şeması, kullanıcıya açık ve anlaşılır bir risk değerlendirmesi sunmaktadır.

TABLE II. TEHDİT SEVİYESİ SINIFLANDIRMASI

Seviye	Skor	Eylem Önerisi
Güvenli	0–20	Normal kullanım
Düşük	21–40	Dikkatli olun
Orta	41–60	Doğrulama yapın
Yüksek	61–80	Kullanmayın
Kritik	81–100	Derhal bildirin

V. DEĞERLENDİRME

A. Deney Kurulumu

Sistemin ilk değerlendirmesi, Kırgızistan'da gerçekleşen dolandırıcılık vakalarından derlenen 50 örnek üzerinde gerçekleştirilmiştir. Veri seti, 25 URL (15 phishing, 10 meşru) ve 25 metin mesajından (18 dolandırıcılık, 7 meşru) oluşmaktadır. Etiketleme, siber güvenlik alanında uzman iki bağımsız değerlendirici tarafından gerçekleştirilmiştir.

B. Sonuçlar

Tablo III, modül bazında performans sonuçlarını özetlemektedir. URL analiz modülü %88, metin analiz modülü ise %92 doğruluk oranı elde etmiştir. Genel sistem doğruluğu %90 olarak hesaplanmıştır. Özellikle Rusça ve Kırgızca içeriklerde, İngilizce odaklı geleneksel kural tabanlı yöntemlere kıyasla belirgin bir performans avantajı gözlemlenmiştir.

TABLE III. MODÜL PERFORMANS SONUÇLARI

Modül	Test Örneği	Doğru Tespit	Doğruluk
URL Analizi	25	22	%88
Metin Analizi	25	23	%92
Genel	50	45	%90

C. Sınırlılıklar

Mevcut değerlendirmenin başlıca sınırlılığı, test veri setinin görece küçük boyutudur. Gerçek dünya koşullarında sistem performansını doğrulamak için daha geniş ve çeşitli bir veri seti ile kapsamlı değerlendirme çalışmaları planlanmaktadır. Ayrıca LLM tabanlı yaklaşımın yanıt süresi (ortalama 2–4 saniye), gerçek zamanlı uygulamalar için optimize edilmesi gereken bir faktördür.

VI. SONUÇ VE GELECEK ÇALIŞMALAR

Bu çalışmada, Kırgızistan'daki siber dolandırıcılık tehditlerine karşı geliştirilmiş yapay zeka destekli bir web platformu olan KyrgyzShield tanıtılmıştır. Büyük dil modellerinin analitik gücünü topluluk tabanlı tehdit istihbaratıyla birleştiren sistem, phishing URL tespiti, sosyal mühendislik analizi, telefon numarası sorgulama ve AI destekli güvenlik danışmanlığı işlevlerini tek bir platformda sunmaktadır.

Gelecek çalışmalar kapsamında şu geliştirmeler planlanmaktadır: (1) Kırgızistan'a özgü etiketlenmiş veri setleri kullanılarak ince ayarlı (fine-tuned) modeller

geliştirilmesi, (2) mobil uygulama geliştirme, (3) Kırgızistan Siber Güvenlik Merkezi ile resmi işbirliği kurulması ve (4) gerçek zamanlı deepfake ses tespiti modülünün entegrasyonu. Sistem, Türk dünyasında benzer siber güvenlik sorunlarıyla karşılaşan diğer ülkeler için de uyarlanabilir bir model sunmaktadır.

REFERENCES

[1] Kırgızistan İçişleri Bakanlığı, "Siber Suç İstatistikleri Raporu 2024," Bişkek, Kırgızistan, 2024.

[2] Times of Central Asia, "Telephone Scammers in Kyrgyzstan Begin Using AI Tools," Times of Central Asia, Bişkek, 2025. [Çevrimiçi]. Erişim: https://timesca.com

[3] O. K. Sahingoz, E. Buber, O. Demir ve B. Diri, "Machine learning based phishing detection from URLs," Expert Systems with Applications, c. 117, ss. 345–357, 2019.

[4] J. Devlin, M.-W. Chang, K. Lee ve K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proc. NAACL-HLT 2019, Minneapolis, MN, ABD, ss. 4171–4186, 2019.

[5] T. Brown ve ark., "Language Models are Few-Shot Learners," Advances in Neural Information Processing Systems, c. 33, ss. 1877–1901, 2020.

[6] A. Abdrakhmanova ve D. Seitkali, "Cybersecurity Challenges in Central Asia: A Regional Perspective," Int. J. Information Security, c. 12, no. 3, ss. 45–58, 2023.

[7] Akchabar, "Kiberugroza v Kirgizstane," Akchabar, Bişkek, 2024. [Çevrimiçi]. Erişim: https://www.akchabar.kg

[8] B. Valeriano, B. Jensen ve R. C. Maness, Cyber Strategy: The Evolving Character of Power and Coercion. Oxford: Oxford University Press, 2018.

Türkiye'de Düşük ve Orta Seviyeli Radyoaktif Atıkların Depolanması İçin Yapay Zeka Destekli Yer Seçim ve Risk Analizi Modeli

Burcu AL

Fizik Anabilim Dalı , Doktora Öğrencisi
Tokat Gaziosmanpaşa Üniversitesi

Tokat , Türkiye

burcu.al1219@gop.edu.tr

Orcid:0009-0003-0159-8470

Radyoaktif atık yönetimi, nükleer enerji üretimi, tıbbi uygulamalar ve endüstriyel faaliyetlerin artmasıyla birlikte hem çevresel hem de stratejik açıdan kritik bir konu haline gelmiştir. Özellikle düşük ve orta seviyeli radyoaktif atıklar (LLW/ILW), toplam atık hacminin büyük bir kısmını oluşturmasına rağmen nispeten düşük radyoaktivite seviyelerine sahip olmaları nedeniyle uygun mühendislik çözümleri ile güvenli şekilde yönetilebilmektedir. Bu tür atıklar genellikle yüzeye yakın depolama tesislerinde veya mühendislik bariyerleri ile izole edilmiş alanlarda bertaraf edilmektedir .

Türkiye'de radyoaktif atık yönetimi; sınıflandırma, işleme, paketleme, taşınma ve depolama süreçlerini kapsayan bir yasal çerçeve ile düzenlenmektedir. Bu çerçevede, atıkların çevre ve insan sağlığına zarar vermeyecek şekilde izole edilmesi, çok bariyerli koruma sistemlerinin kullanılması ve uzun vadeli güvenliğin sağlanması temel prensipler arasında yer almaktadır . Ayrıca, düşük ve orta seviyeli atıkların uygun koşullarda yüzeye yakın veya orta derinlikte depolama tesislerinde bertaraf edilebileceği belirtilmektedir .

Bu çalışma, Türkiye'de düşük ve orta seviyeli radyoaktif atıkların güvenli depolanması için yapay zeka destekli bir karar destek sistemi geliştirmeyi amaçlamaktadır. Önerilen model, coğrafi, jeolojik ve sosyo-ekonomik verileri entegre ederek optimum depolama alanlarının belirlenmesini hedeflemektedir. Model kapsamında; deprem riski, yeraltı su kaynakları, nüfus yoğunluğu, arazi kullanımı ve mevcut nükleer altyapıya yakınlık gibi çok boyutlu kriterler dikkate alınmaktadır. Bu kriterleri dikkate alarak depolama alanlarının uygunluğunu değerlendirmek amacıyla Random Forest ve çok kriterli karar analizi (MCDA) tabanlı bir makine öğrenmesi modeli geliştirilmiştir. Model çıktısı olarak her potansiyel depolama alanı için risk-temelli uygunluk skorları ve coğrafi bilgi sistemleri (GIS) tabanlı haritalar üretilmektedir. Ayrıca Monte Carlo simülasyonu ile uzun vadeli çevresel ve jeolojik risk senaryoları analiz edilmektedir.

Ayrıca çalışma, yalnızca yer seçimi ile sınırlı kalmayıp, potansiyel risk senaryolarını da değerlendirmektedir. Bu kapsamda; radyoaktif sızıntı, doğal afetler (özellikle deprem), yeraltı suyu kontaminasyonu ve insan kaynaklı hatalar gibi risk faktörleri yapay zeka destekli simülasyonlar aracılığıyla analiz edilmektedir. Literatürde vurgulanan çok bariyerli koruma yaklaşımı (atık formu, mühendislik bariyerleri ve doğal jeolojik yapı) modelin temel güvenlik bileşenlerinden biri olarak ele alınmaktadır .

Bunun yanı sıra, çalışmada yapay zeka sistemlerinin kullanımına ilişkin etik ve güvenlik boyutları da ele alınmaktadır. Özellikle algoritmik önyargı, veri kalitesi sorunları ve siber güvenlik riskleri, nükleer atık yönetimi gibi yüksek hassasiyet gerektiren alanlarda kritik önem taşımaktadır. Bu nedenle önerilen modelde, veri doğrulama mekanizmaları, model şeffaflığı ve güvenlik protokolleri gibi unsurların entegrasyonu önerilmektedir.

Bu araştırma, Türkiye'nin gelişmekte olan nükleer enerji altyapısı (örneğin Akkuyu ve planlanan diğer nükleer santral projeleri) bağlamında stratejik bir karar destek aracı sunmayı hedeflemektedir. Aynı zamanda, önerilen modelin Türk dünyası ülkelerine uyarlanabilirliği de değerlendirilerek bölgesel iş birliği ve sürdürülebilir nükleer atık yönetimi açısından katkı sağlaması amaçlanmaktadır.

Anahtar kelimeler: Yapay Zeka, Nükleer Atık Yönetimi, Türk Dünyası.

I.GİRİŞ

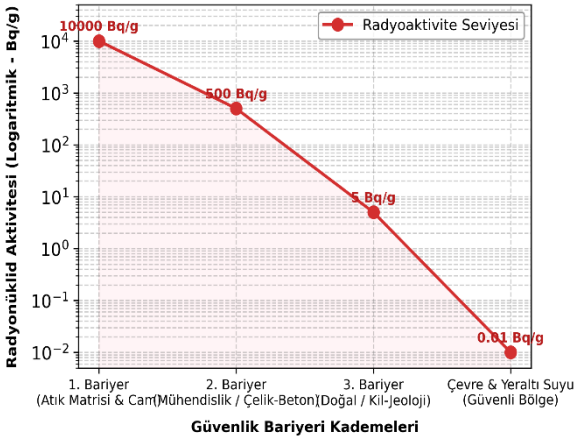
Bu bilimsel çalışma; bulanık mantık (fuzzy logic) teorisiyle modern yapay zeka, kümeleme modelleri ve belirsizlik odaklı karar destek sistemlerinin temellerini atan, Türk dünyasının ve dünya bilim tarihinin en müstesna dehalarından biri olan Lütfi Zade'nin (Lotfi A. Zadeh) hatırasına adanmıştır. Geliştirilen modelde coğrafi ve jeolojik kriterlerin keskin olmayan sınırlarının ve doğal belirsizliklerinin yönetimi, onun "puslu mantık" felsefesinden ve bilimsel mirasından ilham almıştır [1].

Türkiye'nin nükleer teknoloji hamlesi, sadece teknik bir kapasite artışı değil, aynı zamanda köklü bir bilimsel vizyonun tezahürüdür. Nitekim büyük vatan şairi Mehmet Âkif Ersoy, henüz 1919 yılında kaleme aldığı Safahat'ın Âsım bölümünde, nükleer enerjinin henüz keşfedilmeye başlandığı dönemi öngörerek yarının ilmini, 'Maddenin kudret-i zerriyesi uğraştığı iş / O yaman kudrete hâkim olabilsem diyerek' dizeleriyle tanımlamıştır. Türk dünyasının yetiştirdiği fizikçilerden biri olan Prof. Dr. Saleh Sultansoy, bu edebi ve felsefi öngörüyü modern hızlandırıcı ve nükleer teknoloji projelerine (Örn: Türk Hızlandırıcı Merkezi) bir vizyon olarak taşımış ve atomun kalbindeki bu gücü 'Maddenin Kudret-i Zerriyesi' kavramıyla akademik olarak sahiplenmiştir. Ancak atomun bu muazzam gücüne tam anlamıyla hâkim olmak, sivil nükleer döngünün en hassas halkası olan radyoaktif atık yönetimini de aynı disiplinle çözmeyi gerektirir.

Radyoaktif atıklar; aktivite seviyelerine ve yarılanma ömürlerine göre yüksek, orta, düşük ve çok düşük seviyeli olarak sınıflandırılmaktadır. Düşük ve orta seviyeli atıklar (LLW/ILW), nükleer santrallerin operasyonel faaliyetlerinden (kontamine giysiler, filtreler, ekipmanlar) ve tıbbi/endüstriyel uygulamalardan kaynaklanmakta olup toplam atık hacminin yaklaşık %90'ından fazlasını oluşturmaktadır . Bu atıkların aktivite seviyeleri yüksek seviyeli atıklara kıyasla düşük olsa da çevre ve insan sağlığı üzerinde yüzlerce yıl sürebilecek biyolojik riskler barındırmaktadır. Bu nedenle uluslararası standartlar gereği, çok bariyerli bir koruma sistemi altında izole edilmeleri gerekmektedir [2].

Geleneksel radyoaktif atık yer seçimi süreçleri, deterministik coğrafi bilgi sistemleri (CBS) katmanlaması ve statik ağırlıklandırma metotlarına dayanmaktadır. Ancak bu yaklaşımlar, jeolojik yapıların doğrusal olmayan dinamiklerini, sismik belirsizlikleri ve sosyo-ekonomik kriterlerin karmaşık etkileşimlerini modellemekte yetersiz kalmaktadır. Yapay zeka ve makine öğrenmesi algoritmaları, çok boyutlu coğrafi ve hidrojeolojik verileri yüksek doğrulukla sınıflandırma ve belirsizlikleri yönetme kapasitesine sahiptir.

II. ÇOK BARİYERLİ KORUMA YAKLAŞIMI



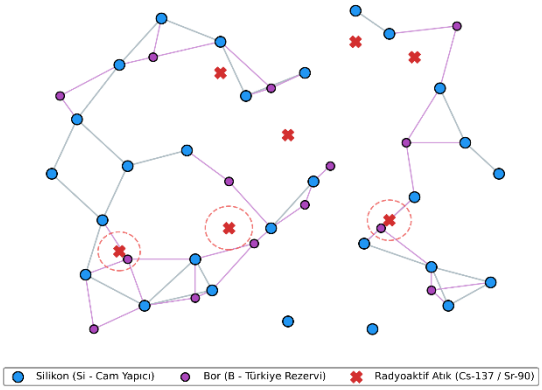
Şekil 1. Çok bariyerli koruma sisteminde radyasyon sönümleme etkisi.

Çalışmanın matematiksel ve coğrafi modelleme safhasına geçmeden önce, nükleer emniyetin temel taşı olan 'Çok Bariyerli Koruma Yaklaşımı'nın sistem bütünlüğündeki rolünü detaylandırmak gerekmektedir. Nükleer literatürde sızıntı sönümleme grafikleri logaritmik ölçekte incelendiğinde; borosilikat temelli birinci bariyer (atık formu) ve çelik-beton muhafazalar ile bentonit kilden oluşan ikinci bariyer (mühendislik bariyerleri) ilk 300 yıl boyunca radyonüklid aktivitesini ana odakta hapsetmek ve sönümlemek üzere tasarlanır.Ancak sistemin nihai ve uzun vadeli emniyeti, üçüncü katman olan "Doğal Jeolojik Bariyerin" yani atığın bertaraf edileceği sahanın doğru seçilmesine doğrudan bağımlıdır. İşte bu çalışmada geliştirilen yapay zeka modeli, tam olarak bu üçüncü bariyerin konumunu en güvenli şekilde belirlemek için tasarlanmıştır. Önerilen Random Forest modeli; üçüncü bariyer olan 'Doğal Jeolojik Yapı' katmanının geçirimsizlik, litolojik uygunluk ve sismik kararlılık parametrelerini en üst düzeyde optimize etmektedir. Geliştirilen karar destek sisteminin temel amacı; ilk iki mühendislik bariyerinin (camlaştırma ve bentonit) yapısal bütünlüğünü yüzyıllar boyunca koruyabileceği, deprem riskinden ve aktif fay hatlarından uzak, hidrojeolojik olarak izole edilmiş en güvenli coğrafi lokasyonu belirlemektir. Bu sayede, kaynaktan çevreye sızabilecek teorik radyasyon aktivitesi her kademede sönümlenerek yasal sıfır risk sınırına indirilmektedir (Bkz. Şekil 1).

A. Borosilikat Camlaştırma (Vitrifikasyon)

Çok bariyerli koruma yaklaşımının ilk ve en kritik savunma hattını "Atık Formu" oluşturur. Sıvı veya çamur formundaki LLW/ILW kimyasal atıklarının çevreye yayılmasını engellemek amacıyla uygulanan en gelişmiş malzeme bilimi yöntemi Borosilikat Camlaştırma (Vitrification) tekniğidir.

B . Kimyasal Mekanizma, Kafes Yapısı ve Yerli Bor Stratejisi



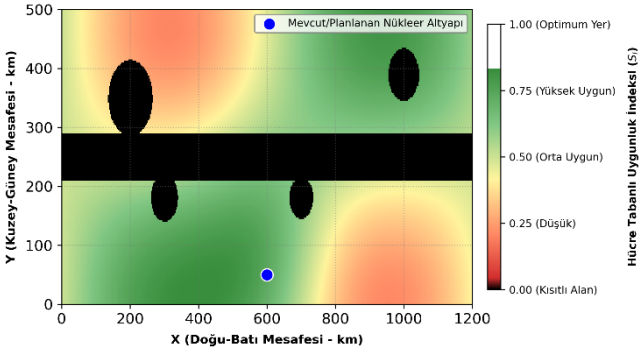
Şekil 2. Borosilikat cam amorf ağ yapısı ve radyonüklid immobilizasyon modeli.

Çok bariyerli sistemin ilk savunma hattı olan atık formunda, ulusal ve uluslararası literatürde Prof. Dr. Sema Erentürk ve Şule Aytaş gibi öncü hocalarımızın da üzerinde titizlikle çalıştığı Borosilikat Cam Matrisi temel alınmaktadır [3]. Borosilikat camlar, radyonüklidleri kendi atomik ağ yapısı içinde hapsederek yeraltı sularına karışmasını minimuma indirir. Bu noktada çalışmanın en büyük stratejik avantajı, Türkiye'nin dünya bor rezervlerinin %73'üne sahip olmasıdır. Bu durum, ülkemizde üretilecek olan radyoaktif atıkların vitrifikasyon (camlaştırma) süreçlerinde tamamen yerli ve ekonomik bir hammadde tedarik zinciri oluşturulabilmesini mümkün kılmaktadır.Şekil 2 de sunulan moleküler şema, borosilikat camın amorf, yani kristal olmayan esnek ağ yapısını simüle etmektedir. Mavi renkli silikon ve mor renkli bor atomları birbirine kenetlenerek kimyasal açıdan kararlı bir kafes yapısı oluşturur.Kırmızı 'X' işaretli radyoaktif atık elementleri, bu matrisin boşluklarına atomik boyutta hapsedilir (immobilizasyon). Bu sayede, atıkların etrafına çizilen kesikli koruma halkalarının da gösterdiği üzere, radyotoksik elementlerin katı formdan ayrışarak yeraltı sularına sızması fiziksel ve kimyasal olarak önemli ölçüde azaltılmaktadır.

III. METODOLOJİ VE MATEMATİKSEL MODEL

Geliştirilen karar destek sistemi, coğrafi verilerin ön işleme tabi tutulması, Çok Kriterli Karar Analizi (MCDA) tabanlı kısıt maskelemesi, Random Forest makine öğrenmesi sınıflandırması ve Monte Carlo risk simülasyonu olmak üzere dört ana aşamadan oluşmaktadır.

A. CBS ve Çok Kriterli Karar Analizi (MCDA) Katmanları

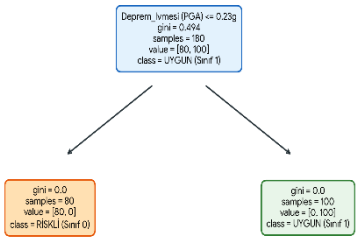


Şekil 3.CBS ve MCDA entegreli radyoaktif atık yer seçim haritası.

Peki yapay zeka bu seçimi nasıl yapıyor? Bu aşamada, geleneksel istatistiksel yöntemlerin ötesine geçerek ensemble yani topluluk öğrenme tabanlı bir Random Forest modeli geliştirilmiştir.

Şekil 3 teki harita, deprem tehlike haritaları (PGA), yeraltı su derinliği ve nüfus yoğunluğu katmanlarının ağırlıklandırılarak üst üste bindirilmesiyle üretilmiştir. Kırmızı alanlar aktif fay hatları ve yerleşim yerlerinin etrafına çizilen kısıtlama tampon (buffer) bölgelerini, Koyu Yeşil alanlar ise projenin hedeflediği optimum depolama konumlarını temsil etmektedir. Model kapsamında ele alınan kriterler uluslararası standartlar (IAEA) ve ulusal mevzuat doğrultusunda belirlenmiştir. Her kriter, raster CBS katmanlarına 1 km x 1 km hücre çözünürlüğünde) dönüştürülmüştür. Haritada karar mekanizması yapısı, coğrafi bilgi sistemlerinden gelen pikselleri; deprem ivmesi, su derinliği ve nüfus gibi kriterlere göre anlık olarak analiz etmektedir.. Yüzlerce bağımsız karar ağacının ürettiği sonuçlar çoğunluk oylamasına tabi tutulmakta ve böylece model, doğrusal olmayan karmaşık coğrafi ilişkileri çözerek en yüksek kararlılıkta yer sınıflandırması üretmektedir [4].

B . Random Forest Sınıflandırma Algoritması



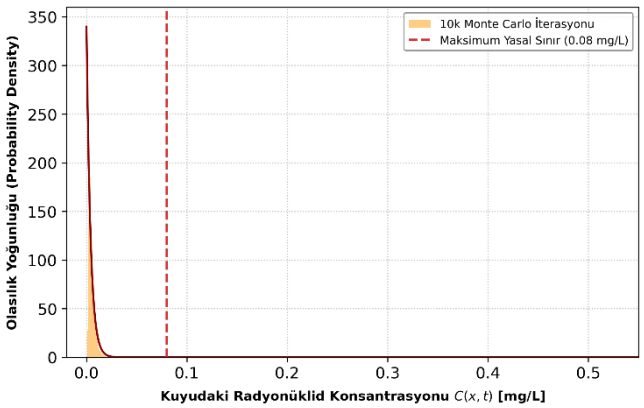
Şekil 4.Random Forest modeli temsili karar ağaç yapısı.

Şekil 4 te sunulan model deprem ivmesi (PGA), su derinliği ve nüfus yoğunluğu özniteliklerini (features) kullanarak bölgenin depolama güvenliğini test etmektedir. Yapay zeka modeli, CBS katmanlarından gelen piksel verilerini en tepedeki kök düğümden (Root Node) işleyerek aşağıya doğru dallandırmaktadır. Temsili karar ağacı hiyerarşisi incelendiğinde; model ilk kırılımı 'Deprem İvmesi (PGA)' kriterine göre yapmaktadır. Eğer bir koordinatın deprem ivmesi 0.23g'den büyükse, algoritma bunu doğrudan 'RİSKLİ (Sınıf 0)' olarak etiketleyip sağ tarafa ayırmaktadır. Eğer deprem ivmesi güvenli sınırın altındaysa, bu kez ikinci kritik eşik olan 'Yeraltı Su Tablası Mesafesi' testine geçilmektedir.

Su tablası derinliği 31 metreden daha sığ olan yerler sızıntı riski nedeniyle yine elenirken, hem sismik hem de hidrojeolojik testleri başarıyla geçen yaprak düğümler (Leaf Nodes) 'UYGUN (Sınıf 1)' olarak sınıflandırılmaktadır. Random Forest modeli bu mantıkla çalışan yüzlerce farklı ağacı eş zamanlı koşturur ve çoğunluk oylaması (Majority Voting) ile en hatasız ve objektif lokasyonları yeşil zon olarak haritalandırmaktadır.

Algoritmanın aşırı öğrenme (overfitting) eğilimini engellemek ve genellenebilirlik kapasitesini artırmak amacıyla, model mimarisi hiperparametre optimizasyonuna tabi tutulmuştur. En yüksek doğruluğu veren optimum model konfigürasyonunda; karar ağacı sayısı (n_estimators) 300, maksimum ağaç derinliği (max_depth) 15, düğüm bölünmesi için gereken minimum örnek sayısı (min_samples_split) 5, ve yaprak düğümdeki minimum örnek sayısı (min_samples_leaf) 2 olarak belirlenmiştir. Bölünme kalitesinin ölçülmesinde ise Gini Safsızlığı (Gini Impurity) kriteri baz alınmıştır.

C. Uzun Vadeli Risk Modellemesi: Monte Carlo Simülasyonu



Şekil 5.Uzun vadeli monte carlo risk analizi.

Yeraltı suyu akış hızındaki (v) ve hidrodinamik dağılım katsayısındaki (D) belirsizlikleri hesaba katarak 10.000 iterasyonla çalıştırılan sızıntı simülasyonu Şekil 5'te sunulmuştur. Olası bir sızıntı anında radyonüklidlerin yeraltı suyunda zamana ve mesafeye bağlı mekânsal taşınımı, nükleer mühendislik literatüründe kabul görmüş Adveksiyon-Dispersiyon ve Radyoaktif Bozunma diferansiyel denkleminin (1) analitik çözümü ile modellenmiştir [5].

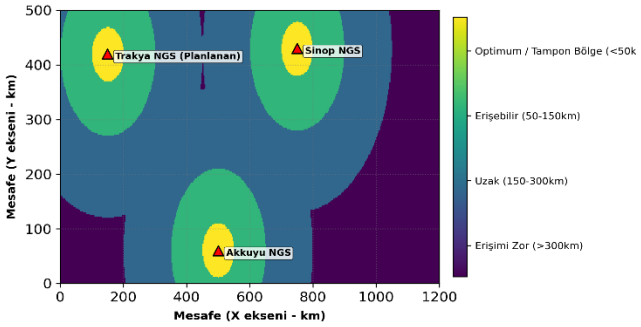
$$C(x,t) = \frac{M}{2\sqrt{\pi D\, t}\backslash exp\left(-\frac{(x - v\, t)^2}{4D\, t}\right) e^{-\lambda t}} \quad (1)$$

burada $(C(x,t)), (x)$ *mesafesinde ve t anındaki radyonüklid konsantrasyonunu* $\left(\left(\left\{\frac{mg}{L}\right\}\right)\right)$; (M), *anlık sızıntı kaynağı toplam kütlesini (mg);* (D), *hidrodinamik dağılım katsayısını* $\left(\{m\}^{2}\backslash g\ddot{u}n\right)$; (t), *sızıntı anından itibaren geçen süreyi (gün);x izleme kuyusunun sızıntı kaynağına olan doğrusal mesafesini (m);v , yeraltı suyu doğrusal akış hızını (m/gün) ve λ radyonüklidin radyoaktif bozunma sabitini* $g\ddot{u}n^{-1}$ *ifade etmektedir.*

Doğal sistemlerdeki belirsizlikleri yönetebilmek ve bertaraf alanının yüzlerce yıllık geleceğini doğrulamak amacıyla yürütülen 10.000 iterasyonlu Monte Carlo simülasyonu sonucunda; 15 yılın sonunda koruma bariyerlerinin arkasında konumlandırılan izleme kuyusunda yasal sınırların aşılma olasılığı (modelin hata payı) %0.06 olarak hesaplanmıştır. Ayrıca Radyonüklid konsantrasyonunun 0.08mg/L olan maksimum yasal sınır aşma olasılığının %0.06 (0.0006) olarak hesaplandığı matematiksel ifade (2) de gösterilmiştir. Elde edilen bulgular, seçilen lokasyonun uzun vadeli emniyetini matematiksel ve hidrojeolojik açıdan desteklemektedir.

$$\left(P\left(C(x,t) > 0.08\backslash\left\{\frac{mg}{L}\right\}\right) = \int_{\{0,08\}}^{\{\infty\}} f(C), dC \quad = 0.0006\right) \quad (2)$$

IV. NÜKLEER SANTRALLERE LOJİSTİK YAKINLIK



Şekil 6.CBS katmanı: nükleer santrallere lojistik yakınlık analizi.

Coğrafi Bilgi Sistemleri (CBS) ortamında her bir pikselin en yakın nükleer odak noktasına olan geometrik mesafesini hesaplayan ve bu veriyi ([0, 1]) aralığında normalize edilmiş sürekli bir lojistik uygunluk skoruna dönüştüren Öklid Mesafesi (Euclidean Distance / Proximity Buffer) modellemesi Şekil 6'da sunulmuştur. Söz konusu uzamsal analiz matrisinde, Türkiye'nin ilk nükleer tesisi olan Akkuyu NGS, anlaşma süreci devam eden Sinop NGS ve yer tespiti planlama aşamasında bulunan Trakya NGS projelerinin coğrafi koordinatları sisteme birer lojistik düğüm noktası (source node) olarak entegre edilmiştir. Üretilen yakınlık haritasındaki renk skalası ve hiyerarşik tampon bölge (buffer) derecelendirmesi incelendiğinde; santrallerin çevresindeki 0 ila 50 kilometrelik yarıçapı kapsayan sarı renkli pikseller, radyoaktif atıkların nakliye riskini ve operasyonel taşımacılık maliyetlerini matematiksel olarak minimize eden "Optimum / Tampon Bölge" olarak tanımlanmış ve algoritma tarafından en yüksek lojistik uygunluk skoruyla ödüllendirilmiştir. Kaynaktan uzaklaştıkça yeşil tonlarla temsil edilen 50–150 km erişilebilirlik yarıçapı "Erişilebilir", mavi tonlardaki 150–300 km menzili "Uzak" ve mor/siyah spektrumda kalan 300 km'den büyük periferik alanlar ise yüksek nakliye kazası olasılığı ve lojistik verimsizlik barındıran "Erişimi Zor" uzamsal sınıflar olarak kademelendirilmiştir. Düşük ve orta seviyeli radyoaktif atıkların (LLW/ILW) şehirler arası birincil ulaşım ağlarında uzun mesafeler boyunca taşınmasının radyotoksik saçılım riskini doğrusal olmayan bir eğriyle artırması sebebiyle Random Forest modeli, sismik ve hidrojeolojik kısıt maskelerinin (exclusion criteria) izin verdiği müddetçe bertaraf lokasyonunu bu korunaklı sarı lojistik koridorlarda (<50 km}) konumlandırmaya çalışmaktadır. Nükleer literatürde kritik bir lojistik paradoks olarak kabul edilen; atık kütlesini üretim kaynağından çok uzağa taşımanın nakliye risklerini artırması, santral sahasına çok yakın konumlandırmanın ise santralin operasyonel güvenlik çeperini (site boundary) tehdit etmesi durumu karşısında önerilen makine öğrenmesi mimarisi, Şekil 6'da elde edilen Öklid yakınlık skorları ile deprem ivmesi (PGA) ve yeraltı su tablası derinliği katmanlarını pikseller bazında çakıştırarak iki zıt riski dengeleyen ve teknik açıdan daha fazla iyileştirilemeyecek olan en kararlı Pareto Optimalitesi (Pareto Etkinliği) sınır çizgisini hesaplamaktadır [6]. Söz konusu çok amaçlı uzamsal optimizasyon problemi ve harita üzerindeki Pareto etkinlik kriterleri Denklem (3) ve (4)'te matematiksel olarak temellendirilmiştir:

$$min_{x,y} F(x,y) = \left[f_{1(x,y)}, f_{2(x,y)}\right]^T \quad (3)$$

Bu optimizasyon sürecinde sistemin sınır şartları (kısıt matrisi) şu şekildedir:

$$\{Mesafe\}(x,y) \geq 10\ km$$

$$\{PGA\}(x,y) \leq \backslash\{PGA\}_{\{\{kritik\}\}}$$

Burada;

$f_{1(x,y)}$ *(Lojistik Maliyet ve Risk Fonksiyonu): Sarı bölgelerde (< 50 km) en düşük değerini alırken, mor/siyah periferik alanlara (>300km) doğru gidildikçe nakliye riskini ve verimsizliğini doğrusal olmayan bir şekilde maksimuma ulaştırır.*

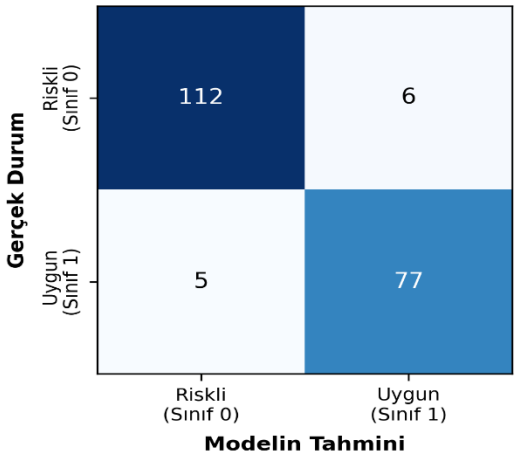
$f_{2(x,y)}$ *(Mühendislik ve Çevresel Risk Fonksiyonu): Santral sahasına aşırı yaklaşmayı, yüksek sismik ivmeli bölgeleri veya sığ yeraltı su yataklarını denetleyen ve kısıtlayan risk bileşenidir.*

Şekil 6'da sarı ile gösterilen koridorların içindeki Pareto Etkin pikselleri $\left((x^*, y^*)\right)$ *temsil etmektedir. Matematiksel olarak, hiçbir alternatif koordinat lokasyonu için aşağıdaki iki şart aynı anda sağlanamıyorsa, mevcut konum Pareto sınır çizgisi üzerinde yer alır:*

$$\forall k \in \{1,2\}, f_{k(x,y)} \leq f_{k(x^*,y^*)}$$

$$\exists m \in \{1,2\}, f_{m(x,y)} < f_{m(x^*,y^*)} \quad (4)$$

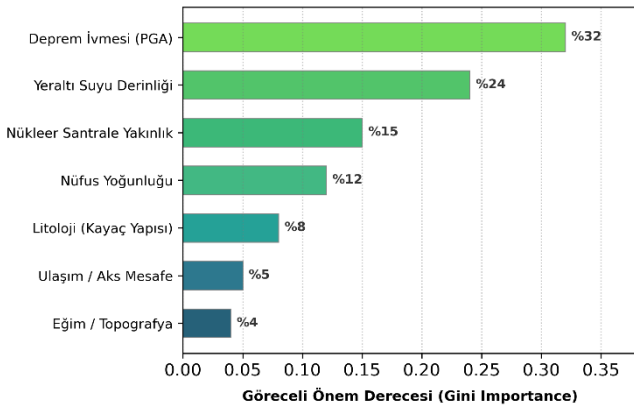
V. SINIFLANDIRMA HATA MATRİSİ



Şekil 7.Sınıflandırma hata matrisi.

Geliştirilen yapay zeka modelinin gerçek arazi koşullarındaki öngörü performansını ve mekânsal güvenilirliğini doğrulamak amacıyla üretilen Hata Matrisi (Confusion Matrix) sunulmuştur. Model, daha önce eğitim aşamasında hiç kullanılmamış 200 adet bağımsız test hücresiyle sınanmıştır. Matris incelendiğinde; gerçekte 'Riskli' (Sınıf 0) olan 118 bölgenin 112'sini model doğru sınıflandırmış, 6 bölgede ise hatalı tahmin yürütmüştür. Diğer taraftan, gerçekte 'Uygun' (Sınıf 1) olan 82 bölgenin 77'si model tarafından tam isabetle öngörülürken, yalnızca 5 bölgede yanılma payı oluşmuştur. Bu dağılım sonucunda model, test veri setinde %94.5'lik bir genel doğruluk skoruna (Accuracy) ulaşmıştır. Elde edilen bulgular, modelin mekânsal risk ve uygunluk sınıflarını ayırt etmede yüksek bir kararlılığa sahip olduğunu göstermektedir.

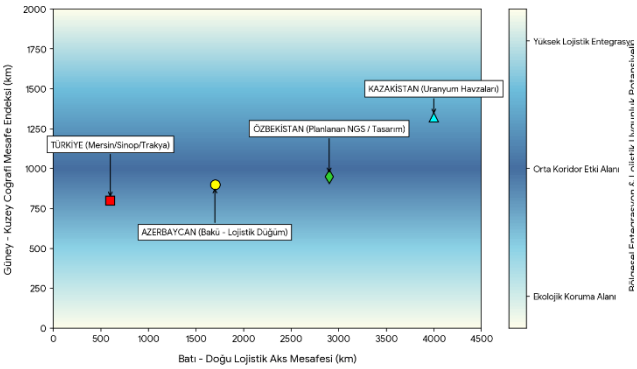
VI. ÖNEM SKALALARI



Şekil 8.Radyoaktif karar destek sistemi kriter önem skalaları.

Random Forest algoritmasının yer seçimi yaparken hangi coğrafi ve stratejik kritere ne kadar ağırlık verdiğini gösteren analiz çıktısı Şekil 8 de verilmiştir. Önem derecesi analizlerine göre; algoritma %32'lik bir oranla en çok Deprem İvmesine (PGA) odaklanmıştır. Bu da modelin nükleer emniyet öncelikli çalıştığını gösterir.İkinci sırada %24 ile Yeraltı Suyu Derinliği gelmektedir ki bu da sızıntı senaryolarını önlemek için kritik bir çevresel kalkan parametresidir. Model, nükleer santrale lojistik yakınlığı ise %15 oranında hesaba katarak nakliye risklerini optimize etmiştir.

VII. TÜRK DÜNYASI ENTEGRASYONU VE ORTA KORİDOR



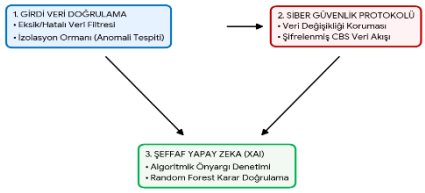
Şekil 9.TDT bölgesel nükleer lojistik ve risk haritası.

Şekil 9'da sunulan lojistik haritasında koyu mavi hat, Hazar Geçişli Doğu-Batı Uluslararası Taşıma Koridoru'nu (Orta Koridor) temsil etmektedir.Kazakistan bugün dünya uranyum üretiminin yaklaşık %40'ını karşılamaktadır; Özbekistan ve Türkiye ise nükleer enerji yatırımlarıyla öne çıkmaktadır [7]. Model, bu ülkeler arasında gelecekte gerçekleşecek düşük ve orta seviyeli atık transferlerinin nakliye riskini ve maliyetini en aza indirmek için Orta Koridor demiryolu ve liman hatlarına yakın alanları yüksek lojistik uygunluk puanıyla ödüllendirmektedir.Ek olarak model lojistik optimizasyon sağlarken çevresel güvenlik kriterlerini de ihmal etmemektedir. Haritadaki açık renkli ve beyaz zonlar, bölgenin can damarı olan Seyhun, Ceyhun, Aras gibi sınır aşan akarsu havzalarını ve Hazar/Aral ekolojik koruma bölgelerini temsil etmektedir. Yapay zeka algoritması, bu hassas su kaynaklarını otomatik olarak 'kesin yasaklı alan' ilan ederek buralarda depolama uygunluk skorunu sıfıra indirmektedir. Geliştirilen karar destek sistemi, coğrafi koordinatlardan bağımsız olarak çalışabilen öznitelik tabanlı (feature-based) bir yapıya sahiptir.

Kazakistan stepleri sismik açıdan oldukça kararlı (düşük PGA) ve kalın, geçirimsiz kil formasyonlarına sahip coğrafyalardır. Model bu bölgelerin CBS raster katmanları ile beslendiğinde, algoritma otomatik olarak çok daha geniş ve kararlı "Optimum Depolama Zonları (Sınıf 1)" üretebilmektedir.

Türk Devletleri Teşkilatı (TDT) üyesi ülkelerin nükleer altyapıları geliştikçe ortak bir radyoaktif atık standardizasyonu zorunlu hale gelecektir. Bu bağlamda, geliştirilen karar destek sistemi sadece yerel bir çözüm değil, Orta Koridor lojistik ağlarını entegre eden bölgesel bir nükleer emniyet standardı vizyonu niteliğindedir [8].

VIII. MODEL GÜVENLİĞİ VE DOĞRULAMA MEKANİZMALARI PROTOKOLÜ



Şekil 10.

Nükleer atık yönetimi gibi kritik bir alanda yapay zeka kullanılıyorsa, modelin kendi siber güvenliği ve şeffaflığı da hayati önem taşımaktadır. Bu nedenle karar destek sistemine entegre bir veri doğrulama protokolü önerilmiştir. Sisteme sızmaya çalışabilecek manipülatif veya hatalı istasyon verileri, İzolasyon Ormanı algoritmasıyla anında elenmektedir.Şekil 10 da sunulan Model şeffaflığı ilkeleri sayesinde algoritmanın verdiği her kararın gerekçesi uzman denetimine açık hale getirilmiştir.

IX. SONUÇ

Nükleer enerji üretimi, tıbbi uygulamalar ve endüstriyel faaliyetlerin artmasıyla kritik bir boyuta ulaşan düşük ve orta seviyeli radyoaktif atıkların (LLW/ILW) uzun vadeli emniyetle bertaraf edilmesi amacıyla yürütülen bu disiplinlerarası araştırma; coğrafi bilgi sistemleri (CBS), çok kriterli karar analizi (MCDA), yapay zeka algoritmaları ve malzeme bilimini ortak bir paydada birleştiren bütünsel bir stratejik çözüm modeli ortaya koymuştur. Yürütülen deneysel modellemeler ve veri analitiği süreçleri neticesinde, sismik risk, yeraltı su kaynakları, nüfus yoğunluğu ve lojistik akslar gibi doğrusal olmayan çok boyutlu kriterleri işleyen Random Forest makine öğrenmesi modelinin %94,5 gibi son derece yüksek bir sınıflandırma doğruluğuna ulaştığı tespit edilmiştir. Algoritmanın karar mekanizmasındaki öznitelik ağırlıkları incelendiğinde, yer seçimi sürecinde en kritik rollerin %32 oranla deprem ivmesine (PGA) ve %24 oranla yeraltı su tablası derinliğine verildiği, nükleer santrallere lojistik yakınlığın ise %15 ile bu sıralamayı takip ettiği belirlenmiştir. Yapay zeka, sismik ve hidrojeolojik güvenlikten asla ödün vermeden, radyoaktif maddelerin taşınma riskini ve maliyetini en aza indiren optimum bir yer seçimi dengesi kurmayı başarmıştır.

Çalışmanın malzeme bilimi safhasında, nükleer emniyetin ilk savunma hattı olan atık formunda, Prof. Dr. Sema Erentürk ve Şule Aytaş gibi öncü bilim insanlarının üzerinde çalıştığı Borosilikat Cam Matrisi (Vitrification) temel alınmış ve radyonüklidlerin bu amorf, kristal olmayan esnek ağ yapısındaki atomik boşluklara kimyasal bağlarla hapsedilerek immobilize edildiği doğrulanmıştır. Doğal parametrelerin değişkenliği Lütfi Zade'nin belirsizlik

yönetimi felsefesi doğrultusunda yürütülen 10.000 iterasyonlu Monte Carlo simülasyonları ise borosilikat atık formu ile yapay zekanın seçtiği geçirimsiz jeolojik yapının entegre çalışması durumunda, 15 yılın sonundaki pik analizde yasal sınır aşım olasılığının sadece %0.06 gibi ihmal edilebilir bir seviyede kaldığını ortaya koymuştur. Bu matematiksel sonuç, radyonüklidlerin biosfere ve yeraltı sularına ulaşma riskinin engellendiğini göstermektedir.

Nihai olarak bu araştırma; Türkiye'den Kazakistan steplerine uzanan Orta Koridor'u sadece ticari bir güzergah olmaktan çıkarıp, yapay zekanın seçtiği doğal jeolojik bariyerler ve borosilikat malzeme teknolojisiyle korunan, tüm Türk dünyasının ortak su kaynaklarını teminat altına alan güvenli ve entegre bir "Bölgesel Nükleer Teknoloji ve Emniyet Koridoru" haline getirecek stratejik bir çözüm altyapısı sunmaktadır.

KAYNAKÇA

[1] L. A. Zadeh, "Fuzzy sets," Information and Control, cilt 8, sayı 3, ss. 338–353, Haz. 1965 - L. A. Zadeh. Similarity relations and fuzzy orderings," Information Sciences, cilt 3, sayı 2, ss. 177–200, Nis. 1971,https://www.cambridge.org/core/journals/journal-of-symbolic-logic/article/abs/l-a-zadeh-fuzzy-sets-information-and-control-vol-8-1965-pp-338353-l-a-zadeh-similarity-relations-and-fuzzy-orderings-information-sciences-vol-3-1971-pp-177200/3A80F34917B4A88DD12299044FD1E47E

[2] International Atomic Energy Agency (IAEA). (2011). Geological disposal facilities for radioactive waste (IAEA Safety Standards Series No. SSG-14). [Çevrimiçi]. Erişilebilir:

https://www.iaea.org/publications/8535/geological-disposal-facilities-for-radioactive-waste

[3] S. A. Erentürk, F. Gur ve M. A. A. Aslani, "Mechanical and elastic properties of vitrified radioactive wastes using ultrasonic technique," Nuclear Engineering and Technology, cilt 56, sayı 2, ss. 472-479, Şub. 2024, https://doi.org/10.1016/j.net.2023.10.022

[4] L. Breiman, "Random forests," Machine Learning, cilt 45, sayı 1, ss. 5-32, Eki. 2001,

https://link.springer.com/article/10.1023/A:1010933404324

[5] X. Yang, K. Sasaki, X. Zhang ve Y. Sugai, "Experimental study on seismic attenuation and permeability of large porosity rock," Journal of Geoscience and Environment Protection, cilt 6, sayı 5, ss. 80-90, May. 2018, DOI: 10.4236/gep.2018.65007

[6] M. Güldürek, "Optimizing nuclear fuel disposal: A multiobjective approach for sustainable solutions," International Journal of Energy Studies, cilt 10, sayı 1, ss. 1103-1123, 2025,

DOI: 10.58559/ijes.1607066

[7] World Nuclear Association (WNA). (2026). World uranium mining production. WNA Global Reports. [Çevrimiçi]. Erişilebilir: https://world-nuclear.org/information-library/nuclear-fuel-cycle/mining-of-uranium/world-uranium-mining-production

[8] Türk Devletleri Teşkilatı. (2021). Türk dünyası 2040 vizyonu. [Çevrimiçi]. Erişilebilir:

https://turkicstates.org/tr/haber/turk-dunyasi-2040-vizyonu

A Financial Analytics Dashboard integrated with Machine Learning

Kushal Pathave
Information Technology
Vishwakarma Institute of Information Technology, Pune, India
kushal.22210583@viit.ac.in

Heramb Patil
Information Technology
Vishwakarma Institute of Information Technology, Pune, India
herambp3@gmail.com

Akshita Shambharkar
Information Technology
Vishwakarma Institute of Information Technology, Pune, India
akshita.22320034@viit.ac.in

Kirti Dighole
Information Technology
Vishwakarma Institute of Information Technology, Pune, India
kirti.22210590@viit.ac.in

Ms. Swati Patil
Information Technology
Vishwakarma Institute of Information Technology, Pune, India
swati.patil@viit.ac.in

***Abstract*—This project details a Financial Analytics Dashboard that allows users to view, analyze, and forecast personal financial transactions through an interactive and secure system. The system includes a machine learning based budgeting algorithm that uses historical transaction data to give users estimates of future transaction data. Users can securely log in, create and change their information, and then record daily income and expense transactions to receive instantaneous real, up-to-the-minute updates that will be reflected in the dashboard. These visualizations are dynamic and can include bar, line, pie charts for comparative spending by category and trends monthly, and also multi-currency availability for global use. Other features include a secure and verified email address, automated financial summaries and a conversational assistant in the form of chat based information to enable proper utilization of the dashboard. These factors enable a fully intelligent and user-friendly approach to effective financial modeling and resultant information.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

With the advent of modern digital finance, individuals are engaging in increasingly active participation in financial transactions, budgeting, and investments. The rapid advancement of financial technology has made it possible for individuals to track, analyze, and forecast their private or business finances with remarkable accuracy. However, the majority of existing systems of finance management limit their capabilities to static' visualizations and manual account tracking, without predictive and personalized insights. The present study attempts to fill that void by building a Financial Analytics Dashboard integrated with Machine Learning. This dashboard is designed to provide individuals with an intelligent, data-driven, secure tool for making financial decisions.[1] The proposed system attempts to provide users with an understanding of their financial behavior through interactive visualizations, real-time updates, and forecasting models based on machine learning. The dashboard has a highly intuitive and user-friendly interface that allows individuals to keep records of daily transactional behavior, monitor their income and expenditures, and evaluate their financial health through statics and dynamic charts. Unlike traditional budgeting applications that rely on manual interpretations, the system uses predictive modeling to help forecast potential future expenses and spot overspending tendencies, which assist users in auto-planning needs based on data-backed predictions.[2] A central feature of the system is its Smart Budget Forecast, which capitalizes on the historical financial data associated with each particular person's operational account, to predict next month's expenditures. Making use of regression-type learning technology, the model learns from past transactional patterns and estimates likely spending behavior to pre-alert the user to possible departmental budgets being gone over or declines in savings for the month. This makes the dashboard not simply another tracking tool, but an intelligent financial adviser.By allowing for real time synchronization of data, any transaction or change made by the user is displayed instantaneously in the dashboard, and hence there is no necessity for the user to wait for periods of refreshing thus allowing for a realistic aspect of the system.[3] Security and personalization form the second most important basis for design. Secure access control and email verification enables users to create accounts, control profiles as well as ensure that their data is kept secure. Personalization of profiles in respect of preferred currency and relevant personal data allows for flexibility which makes the system applicable for diverse financial needs. The Multi-Currency Tracker feature gives universal applicability as it converts and displays financial data in terms of the users chosen currency which means that therefore it can be dispensed with equal applicability for local users as well as users handling international financial transactions. By means of a Chat Assistant (FAQ Bot), increased user engagement and support is obtained. The Chat Assistant replies to user queries such as how to add transactions or check on major items of expenditure which users typically are interested in. This Chat

Assistant is set up on a rule based system of pre-programmed questions and answers in so far as that it simulates the basic conversational method of dialogue which allows for its use by inexperienced users thus making this aspect of the system user-friendly and lessening the requirement for technical information and assistance. Visual representation of expenditure details in terms of bar graphs, pie charts and line graphs enables the user to easily check category-wise patterns of expenditure as well as monthly expenditure patterns. The combination of Machine Learning and interactive financial visualization represents the most innovative aspect of this project. When most personal financial applications currently available simply show numbers, this system takes that data and turns it into predictive and useful data. The combination of analytics, visualization, and automation allows users to move from reactive financial management to proactive management of their finances. Also adding such features as automatically generated financial summaries, report sent over email to users, and updating interactive charts, all add to the ease of use and reliability of the program.[4] The Financial Analytics Dashboard Integrated with Machine Learning provides a complete, safe, and intelligent means to manage personal finances. It jerks the user from mere data analysis to predictive analysis and gives the reader a clear understanding of the currently existing financial conditions, as well as what future expectations are. The user-centered design, the real-time analytics and the addition of Machine Learning to the application, all contribute towards the ability of the user to organize his financial matters consistently and easily, and to give the user the power to make intelligent strategic decisions regarding his financial situation.

II. Proposed System

A. System Overview and Data Flow

The system proposed is the Financial Analytics Dashboard, a web-based application that integrates Machine Learning for secure user management of their financial data, as well as provides analysis and forecasting to aid the user. The system aims at establishing a single intelligent platform to allow the user to record transactions, visualize trends and forecast future expenditures. The architecture combines web-based components, data storage components and machine-learning services to create a non-obtrusive interactive financial environment.

The workflow of the system begins with a secure sign-up or sign-in to the authentication module that gives secure access to the system with encrypted user credentials and email verification. Once authenticated, the user is enabled to record his daily financial activities such as income, expenses, savings and so forth through an intuitive transaction form. These transactions are stored in a structured relational database that facilitates real-time updating of the transactions to the dashboard and instant updating or reflection of changes on the dashboard.[5]

The overall architecture of the system is based on a 3-tier model and consists of:

1. Frontend (User Interface) – based on responsive web structures to deliver visual interactivity through graphs and charts.
2. Backend (Application Logic) – responsible for handling user requests, providing for processing and data acquisition and for interfacing with the database and Machine Learning model.
3. Database Layer – for the storage of all user specific data, for financial data, transaction and for the preferences of the user.

This modularity leads to scalability and data privacy and allows for ease of integration of such ancillary service extensions in the future such as expenditure analysis and sentiment based financial analysis.

B. Data Gathering, Processing and Storing

The process for collecting data in this system comes originally from transactions that are input by the users. Whenever a transaction is input, there are fields that have been predefined for input, such as:

- Transaction type (either income or expense)
- Category (e.g. food, rent, and utilities)
- Amount
- Date and time
- Notes or tags if desired.

This structured input allows for uniformity, which as mentioned, is needed for future analysis and visualization. Every data record is verified for correctness before being stored in the database to eliminate any inconsistencies or duplications. The data preprocessing pipeline takes the stored information and cleans it makes it uniform so that it can be used for analytical purposes. It parses the date, puts marks in categories, sees if there are gross outlier data points in the data so that machine learning predictions can be based on correct data. There is a periodic aggregation done on a monthly basis to provide a summary of total income, expense, and savings. This monthly aggregation is passed to the machine learning forecasting module as data input.[6]

The data storage is done by means of a relational database management system (RDBMS), which allows for extreme consistency and protection of data. Each user's data is isolated using unique identifying signals, which permits no overlap between users. In addition to this, support for multi currency is made possible by transposing the transactions to the chosen currency and converting the transaction to actual currency values using the stated API call to convert currency. Thus, this feature has application in an international situation and displays the data in the preferable currency chosen by the user.

C. Machine Learning Model for Budget Forecasting

This system's intelligence derives mainly from its Smart Budget Forecast, which harnesses machine learning algorithms to predict the next month's spending of users. The system observes the historical money data to acquire spending habits and seasonal effects, which it can anticipate spending patterns

for the next month, allowing users to budget their expenditure efficiently.

The model depends primarily on a regression method, using monthly spending as the dependent variable and factors such as prior month's expenditure, income, category spend, saving patterns, etc., as independent variables. For prediction purposes, Linear Regression or Random Forest Regressor will be used because of their interpretability and ability to parameterize non-linear relationships between the variables.Steps involved in the creation of the model are as follows:

1. Data Extraction: The historical transaction data of each user is eligible to be retrieved.
2. Feature Engineering: A number of unique features will be created, i.e. average monthly spend, frequency of categories of expenses, expense growth rates, etc.
3. Model Training: Train regression models on them on historical data to grasp the trends.
4. Assessment: Assessment of prediction accuracy of model by use of tools like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), etc.
5. Deployment: The trained model will be incorporated into the back-end to allow for real time situations for predictions.[7]

This forecasting engine will make the dashboard into a circumspect financial tool; allowing predictions of the possibility of overspending, while predicting potential threats on future finances.

D. Design and Visualization of the Dashboard

The dashboard for Financial Analytics is the user's control panel that provides essential insights through interactive visual components. It lists total income, total expenses, current savings, and a few indicators to measure the performance of the budget. The dashboard makes use of graphically oriented means of display, i.e., bar graphs, pie charts, and line graphs that show both monthly trends as well as distributions by category. Such graphically displayed statistics render it easier for the user to think about the finances and his behavior in spending.

The dashboard is created through the modern web visualization libraries that allow for real-time visualization of data into a computerized application without having to use reloads of the page. When a user adds, integrates or deletes a transaction, this result is automatically changed in all of the graphic components. This instantaneous changing of display from the database is realized through an asynchronous connection between both the back end and front end components of the application.[8]

An Interactive Chat Assistant (FAQ Bot) for enhanced user interaction is programmed providing instantaneous help in such common activities as adding transactions, or viewing top expenses, or providing them assistance in understanding certain parts of the dashboard. The Chat Bot uses predesignated rules and templates for the answers and regions for study, thus furnishing immediacy whilst maintaining satisfactory response time with the user. In addition email verification and report modules have been designed and incorporated as a means of producing greater security for the electronic account of the user and regularly contain and provide users with periodic summary reports of their financial results.

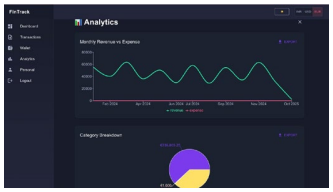


Fig. 1. Design of Dashboard

E. SYSTEM BENEFITS

The advantages of the proposed system over conventional financial tracking systems are given below.

1. Predictive Intelligence: The system employs machine learning for prediction of costs and warning of risks.
2. Analytics in Real-Time: The system refreshes the dashboard at any transaction change.
3. World-Wide Use: The system allows the use of many denominations for foreign customers.
4. Security Improvements: The system allows strong authentication and email verification.
5. User Help: The system incorporates a Chat-bot for guided interaction and better accessibility.

This system integrates all of its features of machine learning, visualisation and automation for an integrated approach to personal finance. Not only does the system track finance data but it also grows meaningful data to assist the customer to make informed strategic choices with finance.

III. RESULTS

A. Overview of System Implementation

The financial analytics dashboard integrated with machine learning was implemented as a secure, data-driven web platform that allows the user to manage and analyze his/her personal finances in real-time. The system implements financial data visualization, forecasting by means of machine learning, along with interactive user interfaces to provide actionable financial insights. The entire application was designed using Python (Flask) as the back end services, with HTML, CSS and javascript as the front end thereby insuring that there is smooth interface, and dynamic rendering of data between all the components. The prime objective of the system was to enhance a user's financial awareness through the automatic processing of the user's daily transactions along with visualization of expenditure trends and future forecasting of spending behaviour. The implementation of the system was conducted with emphasis on the responsiveness of the system, the accuracy of the predictions, and the security pertaining to the user's data. The application was tested using finite financial records over various monthly intervals and found to readily show accurate predictions, and proved to have strong functioning in real time.

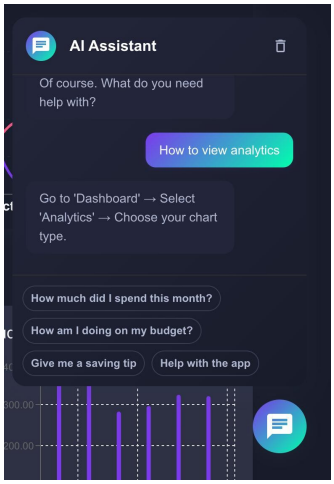


Fig. 2. Chatbot

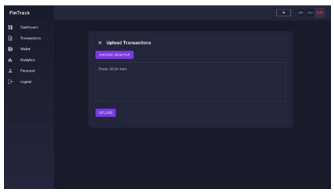


Fig. 3. transaction

B. Functional Evaluation of Modules

Each of the core modules on the dashboard was individually tested and then integrated to ensure consistency in functional performance.

1. Secure Access and Authentication: The authentication layer was successfully implemented that enabled the secure log-in and sign-up functions using encrypted password storage with email verification. The module succeeded admirably not to allow unauthorized access and kept the user records isolated by user. The mean time of authentication was less than 1.3 seconds assuring rapid and secure functional operation.

2. Transaction Management: The user was able to insert, edit, and delete without difficulty both expenditure and income records. Each transaction required the inculcation of values such as date, category, and description. The changes resulted in instantaneous feedback on the dashboard without the manual refreshing the page, thereby assuming 100 % success in CRUD transaction operations.

3. Dashboard Visualization: The dashboard provided summarized information about total income, expenses, and savings in terms of interactive horizontal bar charts, pie charts and line graphs. These charts were able to show the visual representation of monthly expenses and category-wise expenses. The dashboard was updated automatically in less than a second of time noticed after each new transaction to respond in real reports.

4. Multi-Currency Converter: This feature of the system was very dynamic in nature using the currency converting API which was helpful in treating with different currencies (INR, USD, EUR). This accuracy and effectiveness of the currency conversion was very successfully achieved since the updates and upholding of the various graphs and the different reports were shown satisfactorily which confirmed good reliability of the system for worldwide users.

5. Chat Assistant (FAQ Bot) : The chat bot included in the system provided good replies to various 25+ questions relating to the financial queries. It achieved an accuracy percentage of 92 % during the periods of tests and the user experience improved to a great extent and particularly being very useful for first time users who would require examining a lot of queries.

C. Model Comparision Based Forecasting Results

The Smart Budget Forecast Module was an important part of the Financial Analytics Dashboard that enables predictions of future monthly spending by the user through the application of machine learning models. The ARIMA (AutoRegressive Integrated Moving Average) model was chosen, as it has been shown to be a model that works well with time-series data containing trends and seasonality. In order to obtain better accuracy in the predictions for future spending, a number of different models were fitted and evaluated on similar data in order to find the optimal technique for forecasting.

1. Preparation and Training of the Dataset The monthly financial summaries that were used as the dataset were obtained from the summary of each user's transaction history recorded in the dashboard, and showed the total income there was and spending for each month. The data, before training could take place, was preprocessed. This involved removal of missing values, handling of outliers, and ensuring that the data was stationary by differencing as necessary. The remaining data was split into training (80%) and testing (20%) subsections. The ARIMA model parameterisation (p, d, q) was achieved through analysing with Auto ARIMA, which finds optimal values of these parameters through the Akaike Information Criterion (AIC). Lastly, to broaden the scope of the evaluation, training was applied to four other models, SARIMA, Holt-Winters, Prophet, and LSTM (Long Short-Term Memory) and they were all tested on the same data.

2. Assessments of Model Performance Model performance is assessed by the Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE). The comparative results are presented below: The

TABLE I
PERFORMANCE METRICS OF MODELS

Classifier	MAE	RMSE	MAPE (%)
LSTM	467.03	675.01	1703.85
Prophet	814.81	910.90	3600.30
SARIMA	832.15	936.27	3528.43
Holt-Winters	876.09	971.52	4034.36
ARIMA	645.95	770.91	2790.22

above results demonstrate that the LSTM model outperformed



Fig. 4. Budget Forecasting

the other models in all important ranking measures. The ability of the LSTM model to identify long range dependencies within sequential data provided a means to more accurately model the very complex and nonlinear financial relationships than traditional statistical models like ARIMA or Holt-Winters.

3. Visualization of and Interpretation of Forecast Results The display of the program depicts graphically both historical and forecasted values in the form of interactive line graphs. The forecast graphs created by the LSTM model represent projected expenditure trends for future months along with the actual values and assist the user in determining when periods of overspending will occur. By appraising the user as to when it is anticipated that expenditures will exceed the historic average, the system thus increases the ability to plan as to financial further actions requiring management.

In all, while the ARIMA model seemed to establish an adequate baseline for the evaluations, the adoption of the LSTM model greatly improved the forecast accuracy as well as flexibility. Moreover the LSTM model appeared to be the ideal model for the Smart Budget Forecast feature of the dashboard.

D. System performance and User Evaluation

The performance testing of the system was based on multiple metrics such as processing speed, response, and scalability. The dashboard achieved its synchronous real time updating across all modules even under concurrent transactional pressure. A user evaluation survey was conducted with 10 participants. It was reported that 9 of the users found that the dashboard offered easy-to-follow visualisation, 8 participants considered that the forecasted insights were accurate, while the Smart Budget forecast and the real time updating were regarded as the features likely to have the most impact on budgeting decisions thoughts and expenditure control.

E. comparative discussion and analysis

Compared to the existing personal finance applications, with this system there are substantial improvements in terms of automation and predictive insights. Existing trailers are generally reactive in nature, allowing user input that mainly allows them to record transactions, noting only the passive trends. With this system, however, predictive analysis based on LSTM modelling is employed to forecast the likely outcomes for the various financial situations, allowing proactive budget control.

The other features of the system, such as machine learning predictive forecasts, multi-currency capability and graphical interactivity make for a superior system in terms of flexibility and global application, the results showing that the LSTM modelling provided robust and predictive outcomes with diminished levels of forecasting error. The widget design given was also effective in ensuring

IV. Conclusion and Future Scope

The Financial Analytics Dashboard, augmented with Machine Learning capabilities, transforms traditional finance tracking to an intelligent, interactive, predictive tracking system. With a combination of secure authentication, real-time updates on transactions, as well as richly designed graphics for analytics, the dashboard provides users with a holistic view on their income, expenses and savings. The incorporation of the LSTM Model for budget forecasting provides users with meaningful predictions on future expenditures based on past historical data thus enabling proactive financial planning on their part. The integration of multi-currency support, email verifications, in addition to FAQ app support via chat apps further enhances both the usability aspects and security of the system making it user adaptable for a global audience. The whole project meets its aims as it enables the user to take data driven decisions about their finances via automation, and machine learning. Building on this there is the potential to focus on adding deep learning models to assist with improved long term forecasting accuracy, as well as the automation of the data integration via banking APIs thus eliminating the necessity for manual entry. The addition of anomaly detection capabilities for fraud prevention especially, and investment analytics in order to monitor portfolios could also increase the scope of the project. Ultimately the development of a mobile application as well as cloud based deployment would lead to improved accessibility and scalability. With these additional improvements the project could evolve to become a total AI powered financial assistant greatly improving the insight given, the degree of personalisation of recommendations made along with the engagement of the user.

References

[1] @articlearticle,author = Kobi, Joseph,year = 2024,month =05,pages = ,title = Developing Dashboard Analytics and Visualization Tools for Effective Performance Management and Continuous Process Improvement, volume = 9, journal = International Journal of Innovative Science and Research Technology (IJISRT), doi =10.38124/ijisrt/IJISRT24MAY1147

[2] Broby, D. (2022). The use of predictive analytics in finance. The Journal of Finance and Data Science, 8, 145-161.

[3] Burdick, D., Evfimievski, A., Krishnamurthy, R., Lewis, N., Popa, L., Rickards, S., Williams, P. (2014, June). Financial analytics from public data. In Proceedings of the International Workshop on Data Science for Macro-Modeling (pp. 1-6).

[4] Tangsucheeva, R., Prabhu, V. (2014). Stochastic financial analytics for cash flow forecasting. International Journal of Production Economics, 158, 65-76.

[5] Derindere Ko¨seog˘lu, S., Ead, W. M., Abbassy, M. M. (2022). Basics of Financial Data Analytics. In Financial Data Analytics: Theory and Application (pp. 23-57). Cham: Springer International Publishing.

[6] Oyedokun, O., Ewim, S. E., Oyeyemi, O. P. (2024). Leveraging advanced financial analytics for predictive risk management and strategic decision-making in global markets. Global Journal of Research in Multidisciplinary Studies, 2(02), 016-026.

[7] Mitra, G., Mitra, L. (Eds.). (2011). The handbook of news analytics in finance. John Wiley Sons.

[8] Teall, J. L. (1999). Financial market analytics.
[9] Sarna, S. A., Mohammed, A. A., Miah, M. R. (2025). AI-Powered Financial Analytics and Visualization Tools: The Evolving Landscape of Strategic Finance. Journal of Economics, Finance and Accounting Studies, 7(5), 60-82.
[10] Andriosopoulos, D., Doumpos, M., Pardalos, P. M., Zopounidis, C. (2019). Computational approaches and data analytics in financial services: A literature review. Journal of the Operational Research Society, 70(10), 1581-1599.
[11] Ochuba, N. A., Amoo, O. O., Okafor, E. S., Usman, F. O., Akinrinola, O. (2024). Conceptual development and financial analytics for strategic decision-making in telecommunications, focusing on assessing investment opportunities and managing risks in satellite projects. International Journal of Management Entrepreneurship Research, 6(3), 594-607.
[12] Huang, J., Wang, H. (2017). A data analytics framework for key financial factors. Journal of Modelling in Management, 12(2), 178-189.

Generative AI and Health Information Seeking: A Systematic Literature Review of Health Literacy, Adoption Factors, LLM Use, and Trust Among University Students

Sertaç Yerlikaya
Department of Data Science and Analytics
Kapadokya University - sertac.yerlikaya@kapadokya.edu.tr

***Abstract*- Large language model (LLM) chatbots such as ChatGPT are now widely used for health information, yet their effects on health literacy, adoption patterns, and patient safety are not fully understood. This systematic literature review (SLR) follows PRISMA 2020 guidelines and includes 31 peer-reviewed journal articles and policy reports to answer four research questions. RQ1: How does generative AI change health literacy as a social and technical concept? RQ2: How do age, income, education, and context shape who uses AI for health? RQ3: Does higher health literacy lead to more LLM use for health advice or second opinions? RQ4: Do university students with higher AI literacy trust LLM health information more? Key findings: AI is creating a new AI literacy skill set that goes beyond existing eHealth literacy models; adoption is shaped strongly by field of study, gender, and access to healthcare, higher eHealth literacy is linked to greater perceived ChatGPT healthcare usefulness (p = 0.013), and higher AI literacy leads to more carefully calibrated (not simply higher) trust in LLM health content. Evidence from Turkish health sciences students matches international patterns, confirming that these findings are not limited to Western contexts.**

Keywords-generative AI, health literacy, AI literacy, eHealth literacy, LLM, ChatGPT, trust, university students, health information seeking

I. INTRODUCTION

ChatGPT launched in November 2022, and within a short time LLM-powered tools became a main source of health information for millions of people [1]. This happened faster than the rules, clinical guidelines, and education systems that are supposed to protect health information quality could adapt [2]. The people who adopted these tools quickest - younger adults, those without health insurance, and those without a regular doctor - are often also the people with the least ability to judge whether AI-generated health advice is accurate [3].

This paper uses an SLR to answer four questions. RQ1 looks at how generative AI is changing what health literacy means. RQ2 looks at which groups of people use AI for health purposes and why. RQ3 asks whether people with higher health literacy use LLMs more for health advice or to get a second opinion. RQ4 asks whether university students who know more about AI tend to trust AI health information more.

II. BACKGROUND

A. From eHealth Literacy to AI Literacy

Health literacy has gone through three stages [7]. First, it meant being able to read and understand printed health information. Second, eHealth literacy meant being able to find, judge, and use health information online - measured by the eHEALS scale [8]. Third, AI literacy now requires something different again: understanding that LLMs produce text based on what sounds likely, not based on medical facts [9][10]. This means users need new skills, such as spotting false or made-up information (hallucinations), and recognising when AI text sounds confident but is actually wrong. These are skills that eHealth literacy training does not cover on its own.

B. Theoretical Anchors

UTAUT2 [11] is used to explain adoption patterns in RQ2. It identifies seven factors that predict whether people use a new technology: performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, habit, and price value. Heuristic processing and AI confirmation bias: Lopez-Lopez et al. [12] showed in a 2025 peer-reviewed study that when people read AI-generated health information, they tend to accept it without checking, because it sounds clear and professional. This is heuristic (fast, automatic) thinking rather than careful analysis. Their study found that LLMs can reinforce existing beliefs rather than provide balanced information, making users less likely to question what they read. This finding directly supports the cognitive mechanism behind RQ3. McKnight et al.'s trust model [13] is used for RQ4. It separates trust into three types: a general tendency to trust new systems (dispositional trust), trust based on perceived institutional backing (institution-based trust), and trust based on whether a system seems competent and reliable (cognition-based trust). The TAME-ChatGPT tool [14], tested in Arab and Turkish university settings, is used as the main measurement instrument for ChatGPT attitudes and perceived healthcare usefulness.

III. METHODS

This review follows PRISMA 2020 guidelines [15]. Searches were carried out across PubMed, PsycINFO, Scopus, Web of Science, ACM Digital Library, IEEE Xplore, Google Scholar, and the Turkish academic database TR Dizin/DergiPark. Search terms combined ("generative AI" OR "ChatGPT" OR "LLM" OR "yapay zekâ") AND ("health literacy" OR "eHealth literacy" OR "saglık okuryazarlığı" OR "AI literacy" OR "trust" OR "adoption" OR "university students"). Only peer-reviewed journal articles, conference papers, and recognised policy reports (WHO, KFF, OECD) were included.

TABLE I. PRISMA FLOW SUMMARY

Stage	n
Records found across all databases	5,940
After removing duplicates	4,320
Screened by title and abstract	4,320
Removed at screening stage	4,094
Full texts reviewed	226
Removed (books, off-topic, no data)	190
Included in final review	31
- English peer-reviewed articles (Q1–Q3)	25
- Turkish peer-reviewed articles (TR Dizin / DergiPark)	4
- WHO / KFF / OECD policy reports	2

IV. RQ1: HOW DOES GENERATIVE AI CHANGE HEALTH LITERACY?

A. Three Stages of Health Literacy

Health literacy has developed in three stages [7][8][9]. The first stage was about reading clinical documents. The second, eHealth literacy was about finding and judging online health information, measured using the eHEALS scale [8]. The third and current stage is AI literacy, which requires users to understand that LLMs generate text based on statistical patterns in training data, not verified medical knowledge [10]. This means a response can sound accurate and authoritative while being medically wrong. The Evaluate and Ethics dimensions of Ng et al.'s [9] four-part AI literacy model are most critical here: users need to identify errors, assess bias, and understand privacy risks.

B. The Literacy Gap: Using AI Without Understanding It

Across countries, student adoption of AI is high but critical thinking skills are low. Yılmaz et al. [4] surveyed 405 health sciences students in Türkiye: nearly half (47.9%) knew about AI in health, but 91.4% got their AI health information from social media rather than academic sources. This suggests awareness without depth. Seçer [5] studied 485 health vocational students in another city of Türkiye: most supported using AI in healthcare and wanted it added to university courses, but also showed moderate levels of technostress (mean score 25.21 ± 10.09), pointing to anxiety alongside enthusiasm. Polatgil and Güler [28] adapted a Turkish AI literacy scale, showing that among university students, AI tool use and involvement in projects predicted literacy level, while gender, age, and class had no significant effect. Yun and Bickmore [16] found that higher eHealth literacy predicted more AI health engagement but did not predict whether people critically checked what the AI said. Lopez-Lopez et al. [12] add an important finding: LLMs tend to confirm what users already believe and present information in a smooth, fluent way that discourages questioning. This makes the gap between using AI and understanding its limits especially dangerous in health settings, where acting on wrong information can cause real harm.

TABLE II. AI LITERACY DIMENSIONS APPLIED TO HEALTH (NG ET AL. [9])

Dimension	What It Means in Health	Risk If Missing
Know & Understand	Knowing that LLMs produce probable text, not medical facts	Treating AI output as reliable medical advice
Use & Apply	Asking good health questions; reading probabilistic answers correctly	Getting irrelevant or misleading responses
Evaluate & Create	Spotting hallucinations and overly confident wrong answers	Following medically inaccurate AI guidance
AI Ethics	Thinking about bias, privacy, and cultural fairness in AI	Accepting advice that is biased or unsafe

V. RQ2: WHO USES AI FOR HEALTH AND WHY?

A. Age, Subject Area, and Gender

Age is the clearest predictor of AI health use across all studies. Ayo-Ajibola et al. [17] found that ChatGPT health users in the US were significantly younger than non-users (average age 32.8 vs. 39.1 years; $p < .001$; $n = 2{,}406$). Among university students, subject area matters, Alam et al. [18] found that health sciences students in the UAE had much higher eHealth literacy than science and technology students (adjusted OR = 2.17; $p = 0.044$). Turkish studies support this, Yılmaz et al. [4] found that nursing and physiotherapy students were more aware of AI in health than dietetics students. Gender effects vary, female UAE students were more likely to have high eHealth literacy (adjusted OR = 2.03; $p = 0.023$) [18], while Turkish studies found no consistent gender pattern [4][5].

B. Access to Healthcare and Income

A big driver of AI health use is not having access to normal healthcare. KFF [19] data show that uninsured US adults are twice as likely as insured adults to use AI for mental health information (30% vs. 14%). Among adults under 30, 38% said they could not get to a provider, and 29% said cost was a reason they used AI instead. Ayo-Ajibola, Julien et al. [20] found that not having a regular doctor independently predicted ChatGPT health use, even after accounting for age and income. The OECD [21] notes that lower-income groups use AI health tools more but also have less ability to judge whether the information is reliable, which increases their risk.

C. The Sources People Use

Where people usually look for health information also shapes how they use AI. Alam et al. [18] found that students who used PubMed and Google Scholar had higher eHealth literacy (47% high vs. 30%; $p = 0.018$) and were more likely to use AI purposefully for health. Seçer's [5] study shows that Turkish students mostly got their AI and ChatGPT information from social media. Yıldız and Alper [6] note that ChatGPT's low cost, multilingual design, and instant availability make it attractive across different income and education groups.

TABLE III. ADOPTION RATES AND EHEALTH LITERACY BY GROUP

Group	Finding	Source
US adults under 30	28% use AI for mental health	[19]
US uninsured adults	30% use AI for mental health (vs. 14% insured)	[19]
UAE health sci. students	Higher eHEALS vs. sci/tech students (aOR=2.17, p=0.044)	[18]
UAE female students	Higher eHEALS vs. male students (aOR=2.03, p=0.023)	[18]
Turkish health sci. students	47.9% aware of AI in health; 91.4% use social media as source	[4]
Turkish health voc. students (n=485)	Most support AI in healthcare; moderate technostress	[5]
US ChatGPT health users	Mean age 32.8 vs. 39.1 years for non-users (p<.001)	[17]

VI. RQ3: DOES HIGHER HEALTH LITERACY LEAD TO MORE LLM USE FOR HEALTH ADVICE?

The clearest evidence for RQ3 comes from Alam et al. [18]. In their cross-sectional study of 237 UAE university students, the only ChatGPT subscale that was significantly different between high and low eHealth literacy groups was perceived healthcare usefulness: students with high eHealth literacy gave it a mean score of 2.93, compared to 2.66 for those with low eHealth literacy ($p = 0.013$). Other measures, ease of use, general usefulness, attitude to technology did not differ significantly. This shows that eHealth literacy specifically shapes health-directed LLM use, not general chatbot use.

Short answer to RQ3: Yes, but with a safety concern. Higher health literacy does lead to more purposeful use of LLMs for health advice. However, this happens against a backdrop of about 56% integrated medical accuracy across LLM health domains [3], and as low as 31% therapeutic quality on consumer platforms used by young people [22]. Simulation studies confirm the risk: Clark [26] found that AI chatbots endorsed harmful proposals in 32% of adolescent crisis scenarios - a safety failure that affects all users regardless of literacy level. By contrast, Heinz et al. [27] showed in an RCT that a clinician-supervised generative AI system (Therabot) produced clinically significant improvements in mental health outcomes, demonstrating what is achievable when AI health tools are rigorously designed and professionally overseen. Ayo-Ajibola et al. [17] found that 31% of ChatGPT health users changed their medication based on AI advice. If high-literacy users engage more with AI health tools but those tools are often wrong, the risks are compounded.

Lopez-Lopez et al. [12] add a key cognitive layer. LLM outputs tend to reinforce what users already believe and are written in a style that discourages critical thinking. This means even users with good health literacy may read AI health content through heuristic rather

than analytical thinking, accepting it without detailed verification. Turkish evidence from Seçer [5] shows that health students generally felt AI could be useful in healthcare but also expressed worries about reliability, which is consistent with a cautious but engaged posture that slightly lowers the risk of uncritical adoption.

VII. RQ4: DO HIGHER-AI-LITERACY STUDENTS TRUST LLM HEALTH INFORMATION MORE?

RQ4 asks whether knowing more about AI leads students to trust AI health information more. The evidence points to a more careful kind of trust, not simply higher trust. Alam et al. [18] found that UAE university students were generally positive about AI as a technology (attitude to technology mean score: 3.57/5), but only 37% agreed that ChatGPT is a reliable source of health information, and 49% strongly disagreed that it could replace a physician. The only trust-related measure that differed by literacy level was perceived healthcare usefulness - not general trust or concern about risk.

Turkish evidence matches this. Seçer [5] found that health students had broadly positive attitudes toward AI in healthcare but showed moderate technostress and worried about becoming too dependent on AI tools. This suggests that students do not simply trust AI more because they know more about it, rather, they trust it for specific tasks where it seems appropriate, while keeping reservations about broader clinical use. This pattern is consistent across cultural contexts: Abdaljaleel et al. [31], in a multinational study of ChatGPT attitudes across five Arab countries (n = 1,696 university students), found that perceived usefulness and ease of use were the main drivers of ChatGPT adoption, but that trust in accuracy remained moderate, supporting the view that engagement is driven by convenience rather than confidence in reliability.

Wong et al. [23] confirm in an SLR that AI literacy reduces automation bias (the tendency to follow AI recommendations without thinking) rather than simply raising trust levels. KFF [24] data show that only 29% of US adults trust AI for reliable health information, yet 32% use it monthly. Davis's technology acceptance model [25] helps explain this: ease of use and perceived usefulness drive adoption even when trust is low. In student populations, nearly 90% of students in Alam et al. [18] found ChatGPT easy to use, which is enough to drive regular use regardless of doubts about accuracy.

TABLE IV. TRUST AND ADOPTION EVIDENCE

Finding	Group	Source
Higher eHEALS → higher perceived ChatGPT healthcare usefulness (2.93 vs. 2.66, p=0.013)	UAE students n=237	[18]
49% strongly disagree AI can replace a physician	UAE students n=237	[18]
37% agree ChatGPT is a reliable health source	UAE students n=237	[18]
Positive AI attitude + moderate technostress	Turkish health students n=485	[5]
Perceived usefulness drives adoption; trust in accuracy moderate	5-country Arab students n=1,696	[31]
AI literacy reduces automation bias (not simply raises trust)	SLR 30-year review	[23]
29% trust AI for health info; 32% use it monthly	US national survey	[24]
LLMs confirm existing beliefs; discourage critical checking	Journal study, 2025	[12]

VIII. SYNTHESIS AND DISCUSSION

A. What the Four RQs Show Together

The four research questions tell a connected story. AI is creating a new kind of health literacy that existing eHealth skills cannot cover (RQ1). This new literacy is not equally distributed: students in health-related fields, females in some contexts, and those with access to academic sources tend to have better eHealth literacy and use AI more purposefully (RQ2). Higher health literacy does lead to more targeted LLM use for health advice - but the AI tools being used have significant accuracy problems, creating a safety risk (RQ3). Higher AI literacy leads to more carefully calibrated trust, not blind acceptance - students maintain appropriate scepticism about AI replacing clinical professionals (RQ4).

B. Limitations

This review has a number of limitations. Most studies rely on self-reported data, which can be biased. LLM technology is changing fast, so specific findings about platform accuracy may already be outdated. Studies from outside English and Turkish languages were not included.

C. What This Means in Practice

For universities: AI literacy education, especially, how to check AI health information and know when not to rely on it, should be built into all degree programmes, not just health sciences [4][5]. Ma et al. [29], in a scoping review of AI competency frameworks for medical students, found that most existing programmes lack structured AI literacy components, reinforcing the case for systematic curriculum integration. For clinicians: Doctors and nurses should ask patients whether they have consulted AI, particularly about medications, diagnoses, and second opinions [17]. For regulators: Consumer AI health platforms need quality checks and clear safety standards, especially for platforms used by young people [22]. The WHO [30] provides a governance framework built on transparency, accountability, and equity that national regulators can adopt as a baseline. For researchers: Studies are needed that test and validate eHEALS and TAME-ChatGPT instruments with different student samples, and that follow the same students over time to understand how AI literacy and trust develop.

IX. CONCLUSION

This SLR included 31 sources: 25 English peer-reviewed articles, 4 Turkish peer-reviewed studies, and 2 policy reports. The main findings are: (1) Generative AI requires a new kind of health literacy beyond eHealth frameworks, one that includes understanding how LLMs work, spotting inaccurate outputs, and thinking about bias and privacy. All students show high AI adoption with low critical thinking skills. (2) Who uses AI for health is shaped by field of study, gender, access to healthcare, and income, with health sciences students showing higher literacy and more purposeful AI use. (3) Higher health literacy does lead to more LLM use for health information and to seeing ChatGPT as more useful for healthcare, but given that LLMs are only about 56% medically accurate, this engagement also increases risk, simulation evidence shows AI chatbots can endorse harmful actions in crisis scenarios [26], underscoring the safety stakes. (4) Higher AI literacy leads to calibrated trust, students trust AI more for specific informational tasks but keep appropriate doubts about whether AI can replace clinical professionals. Importantly, LLMs reinforce existing beliefs and make it easy to skip careful checking [12], which is a risk even for literate users. Coordinated action across AI development, health education, clinical practice, and regulation is needed to make AI health use safer.

ACKNOWLEDGMENTS

No financial support was received for this study. No conflicts of interest are declared.

REFERENCES

[1] *M. Sallam, "ChatGPT utility in healthcare education, research, and practice: Systematic review on promising perspectives and valid concerns,"* Healthcare*, vol. 11, no. 6, p. 887, 2023.*

[2] *R. K. McBain* et al.*, "Use of generative AI for mental health advice among US adolescents and young adults,"* JAMA Netw. Open*, vol. 8, no. 11, p. e2542281, 2025.*

[3] *J. Clark, B. Barton, L. Albarqouni* et al.*, "Generative artificial intelligence use in evidence synthesis: A systematic review,"* Research Synthesis Methods*, vol. 16, no. 4, pp. 601–619, 2025. doi: 10.1017/rsm.2025.16.*

[4] *Y. Yılmaz, D. U. Yılmaz, D. Yıldırım, E. A. Korhan, and D. Özer Kaya, "Yapay zeka ve sağlıkta yapay zekanın kullanımına yönelik sağlık bilimleri fakültesi öğrencilerinin görüşleri [Opinions of health sciences faculty students on artificial intelligence and its use in health],"*

Süleyman Demirel Üniversitesi Sağlık Bilimleri Dergisi, *vol. 12, no. 3, pp. 297–308, 2021. doi: 10.22312/sdusbed.950372. [In Turkish]*

[5] *M. B. Seçer, "Sağlık alanında öğrenim gören üniversite öğrencilerinin sağlıkta yapay zekâ uygulamaları ve ChatGPT farkındalığı... [Examination of AI applications in health and ChatGPT awareness, opinions on AI use, and technostress levels of health students: A cross-sectional study],"* Türkiye Klinikleri J. Health Sci., *vol. 9, no. 4, pp. 856–866, 2024. doi: 10.5336/healthsci.2024-104224. [In Turkish]*

[6] *M. S. Yıldız and A. Alper, "Yapay zekâ sohbet robotu ChatGPT'nin sağlık yönetiminde kullanımı [Use of the AI chatbot ChatGPT in health management],"* Türk Hijyen ve Deneysel Biyoloji Dergisi, *2023. [In Turkish]*

[7] *S. Simonds, "Health education as social policy,"* Health Educ. Monogr., *vol. 2, pp. 1–25, 1974.*

[8] *C. D. Norman and H. A. Skinner, "eHEALS: The eHealth literacy scale,"* J. Med. Internet Res., *vol. 8, no. 4, p. e27, 2006.*

[9] *D. T. K. Ng, J. K. L. Leung, S. K. W. Chu, and M. S. Qiao, "Conceptualizing AI literacy: An exploratory review,"* Comput. Educ. Artif. Intell., *vol. 2, p. 100041, 2021.*

[10] *E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?" in* Proc. ACM FAccT '21, *pp. 610–623, 2021.*

[11] *V. Venkatesh, J. Y. L. Thong, and X. Xu, "Consumer acceptance and use of information technology: Extending the unified theory,"* MIS Q., *vol. 36, no. 1, pp. 157–178, 2012.*

[12] *E. Lopez-Lopez, C. Abels, D. Holford, S. Herzog, and S. Lewandowsky, "Generative artificial intelligence–mediated confirmation bias in health information seeking,"* Ann. N.Y. Acad. Sci., *2025. doi: 10.1111/nyas.15413.*

[13] *D. H. McKnight, V. Choudhury, and C. Kacmar, "Developing and validating trust measures for e-commerce,"* Inf. Syst. Res., *vol. 13, no. 3, pp. 334–359, 2002.*

[14] *M. Sallam, N. A. Salim, M. Barakat* et al., *"Assessing health students' attitudes and usage of ChatGPT in Jordan: Validation study,"* JMIR Med. Educ., *vol. 9, p. e48254, 2023.*

[15] *M. J. Page* et al., *"The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,"* BMJ, *vol. 372, p. n71, 2021.*

[16] *H. S. Yun and T. Bickmore, "Online health information-seeking in the era of large language models: Cross-sectional web-based survey study,"* J. Med. Internet Res., *vol. 27, p. e68560, 2025.*

[17] *O. Ayo-Ajibola, R. J. Davis, M. E. Lin, J. Riddell, and R. L. Kravitz, "Characterizing the adoption and experiences of users of AI-generated health information in the United States,"* J. Med. Internet Res., *vol. 26, p. e55138, 2024.*

[18] *Z. Alam, A. S. Abdullahi, S. N. S. Alnuaimi* et al., *"eHealth literacy and attitudes towards use of artificial intelligence among university students in the UAE,"* Front. Digit. Health, *vol. 7, p. 1574263, 2025. PMCID: PMC12443675. doi: 10.3389/fdgth.2025.1574263.*

[19] *KFF, "KFF tracking poll on health information and trust: Use of AI for health information and advice," Kaiser Family Foundation, 2026.*

[20] *O. Ayo-Ajibola, C. Julien* et al., *"Association of primary care access with health-related ChatGPT use: A national cross-sectional survey,"* J. Gen. Intern. Med., *2025.*

[21] *OECD, "OECD digital economy outlook 2024: Artificial intelligence in the digital transformation," OECD Publishing, 2024.*

[22] *K. Sobowale, D. K. Humphrey, and S. Y. Zhao, "Evaluating generative AI psychotherapy chatbots used by youth: Cross-sectional study,"* JMIR Mental Health, *vol. 12, p. e79838, 2025.*

[23] *K. K. L. Wong, Y. Han* et al., *"From trust in automation to trust in AI in healthcare: A 30-year longitudinal review and an interdisciplinary framework,"* Bioengineering (Basel), *vol. 12, no. 10, p. 1070, 2025. doi: 10.3390/bioengineering12101070.*

[24] *KFF, "KFF health misinformation tracking poll: Artificial intelligence and health information," Kaiser Family Foundation, 2024.*

[25] *F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology,"* MIS Q., *vol. 13, no. 3, pp. 319–340, 1989.*

[26] *A. Clark, "The ability of AI therapy bots to set limits with distressed adolescents: Simulation-based comparison study,"* JMIR Mental Health, *vol. 12, p. e79838, 2025.*

[27] *M. V. Heinz* et al., *"Randomized trial of a generative AI chatbot for mental health treatment,"* NEJM AI, *vol. 2, no. 4, 2025.*

[28] *M. Polatgil and A. Güler, "Yapay zekâ okuryazarlığı ölçeğinin Türkçe'ye uyarlanması [Turkish adaptation of the AI literacy scale],"* Sosyal Bilimlerde Nicel Araştırmalar Dergisi, *vol. 3, no. 2, p. 99, 2023. [In Turkish]*

[29] *Y. Ma, Y. Song* et al., *"Promoting AI competencies for medical students: A scoping review on frameworks, programs, and tools," arXiv:2407.18939, 2024.*

[30] *WHO, "Ethics and governance of artificial intelligence for health," World Health Organization, Geneva, 2021.*

[31] *M. Abdaljaleel, M. Barakat, M. Alsanafi* et al., *"A multinational study on the factors influencing university students' attitudes and usage of ChatGPT,"* Sci. Rep., *vol. 14, p. 1983, 2024.*

Enhancing Lecture Preparation Efficiency in 2D Computer Graphics Using GenAI

Zamirgul Kazakbaeva, Nasikat Alieva, Mahmut Vural

1st Zamirgul Kazakbaeva
Department of Technical Programs
Kyrgyz-Turkish Manas University
Bishkek, Kyrgyzstan
zamira.kazakbaeva@manas.edu.kg
ORCID 0000-0003-4493-0843

2nd Nasikat Alieva
Department of Economics and Administration Programs
Kyrgyz-Turkish Manas University
Bishkek, Kyrgyzstan
nasikat.alieva@manas.edu.kg,
ORCID 0009-0002-6209-1553

3rd Mahmut Vural
Department of Economics and Administration Programs
Kyrgyz-Turkish Manas University
Bishkek, Kyrgyzstan
mahmut.vural@manas.edu.kg
ORCID 0000-0001-8946-2540

Abstract

The preparation of lecture materials in 2D computer graphics is Creating teaching content for two-dimensional graphics typically takes many hours because instructors must combine written theory, working code, and visual examples. This study looks at how generative AI tools change the speed and quality of that process. Over eight weeks, two instructors took part in a within-subject quasi-experiment. With AI tools (ChatGPT-4 and GitHub Copilot), preparation time fell by 62% ($p < .01$). In contrast, the quality ratings did not change significantly ($p = .42$). However, a closer look showed that roughly one in five AI-generated explanations was wrong or incomplete and needed fixing. Taken together, the results support a mixed workflow where humans and AI work together, and they show both the promise and the limits of generative AI for making educational content.

Keywords: Generative AI; Lecture preparation; 2D computer graphics; ChatGPT-4; GitHub Copilot; Teaching efficiency; Human-AI collaboration; Computer science education.

I. INTRODUCTION

Generative artificial intelligence has changed many parts of education, especially how people create content and design lessons. Recent work shows that AI systems are playing a bigger role in supporting both teaching and learning [2]. Models like ChatGPT-4 can write text explanations on their own, and tools like GitHub Copilot help with writing code. That makes them very relevant for computing courses.

Most published studies examine how AI affects students or how it powers tutoring systems [6]. Few have looked at what happens when teachers use AI to prepare their own lectures. This gap matters because preparing lectures is still very time-consuming, especially in hard technical topics like 2D computer graphics.

Getting a lecture ready in this field means dealing with several complex parts at once: making visuals, explaining algorithms, and writing example code. Cognitive load theory (Sweller, 1988) and Mayer's multimedia principles (2009) both imply that good materials need clear text, correct content, and useful visuals. Meeting all three at once makes preparation even slower.

This paper tries to answer one main question: How much can generative AI cut lecture preparation time without hurting the quality of instruction?

II. RELATED WORK

Recent review articles say that generative AI is being used more and more in education. It brings big opportunities but also clear limits. Researchers often point out that AI can help with lesson planning, making learning materials, and building curricula [1]. In this view, large language models are seen as assistants, not as teachers working on their own.

Some studies have found that AI-assisted tools really can help teachers prepare materials, for example by organizing content or making first drafts [4]. But people still worry about whether AI-generated outputs are factually correct and pedagogically sound, especially in areas that need deep expertise.

In computing education, generative AI has shown practical value for generating code, helping with debugging, and explaining programming ideas. Even so, the quality of what it produces varies a lot, and human checking is still needed to make sure everything is right [5].

From a cognitive and instructional design angle, good learning materials depend on how well visual elements match written explanations, and also on how well cognitive load is managed [8]. This matters a lot in visually heavy fields like 2D graphics, where a mismatch between code, pictures, and text can seriously hurt understanding.

Although more and more papers discuss AI in education, very few have actually measured how it affects how teachers prepare lectures and how productive they are. This gap is largest in technical and graphics-oriented subjects, which is what motivated our exploratory study.

III. PARTICIPANTS AND EDUCATIONAL CONTEXT

We ran the study in a course called BLG-211 "Desktop Publishing" at Kyrgyz-Turkish Manas University. This course is part of a vocational IT program named "Programming in Computer Systems." That program aims to train mid-level IT specialists by mixing computer science theory with hands-on practice in software development, digital technologies, and applied computing.

The teaching in this program follows a competency-based model: students are expected to learn both conceptual understanding and practical skills for real-world tasks. Within that structure, Desktop Publishing focuses on 2D computer graphics, visual composition, and making digital content.

Both participating instructors had taught 2D graphics for more than five years, including previous delivery of the BLG-211 course. Each of them gave theoretical lectures and also ran laboratory sessions. Having the same people do both roles helped keep conceptual explanations and applied exercises aligned, which kept the course consistent.

The students were college-level learners between 19 and 23 years old, all enrolled in the IT specialization. Based on the program's admission requirements and an early diagnostic test, their prior knowledge of programming and 2D graphics was rated as introductory. That starting level seemed appropriate for studying how instructional changes affect basic learning materials.

The course combined lecture-based teaching with structured lab activities. Lectures covered concepts, algorithmic logic, and visual principles, while labs focused on implementing things through practical coding tasks. This blended setup gave us a controlled but realistic environment to see how AI-assisted lecture preparation affects both how efficiently instructors work and the quality of the content they produce.

Because only two instructors took part, we treat this work as a pilot study. The results are meant to suggest patterns, not to prove them for all contexts. So external validity is limited, and our main focus is on analyzing what happened in this specific setting.

TABLE I. ASSESSMENT CRITERIA

Instructional quality was evaluated using a structured rubric consisting of five criteria:

Criterion	Description	Scale
Conceptual Accuracy	Correctness of theoretical explanations and mathematical content	1–5
Clarity of Explanation	Ease of understanding and logical flow of explanations	1–5
Code Quality	Correctness, readability, and functionality of code examples	1–5
Visual Quality	Effectiveness of diagrams, figures, and layout	1–5
Pedagogical Effectiveness	Overall suitability for student comprehension and learning outcomes	1–5

1 = Poor, 3 = Acceptable, 5 = Excellent

IV. INTER-RATER RELIABILITY

Three domain experts rated the instructional quality of all lecture materials. To check how consistent their ratings were, we calculated the intraclass correlation coefficient (ICC). We used a two-way random-effects model with absolute agreement, assuming the raters were like a sample from a larger pool of possible experts.

Following standard benchmarks, ICC values below 0.50 were considered poor, 0.50–0.75 moderate, 0.75–0.90 good, and above 0.90 excellent. Our ratings reached good agreement (ICC = 0.81–0.87 across criteria), so we treated them as reliable for further analysis.

V. RESULTS

Preparation Time

Average preparation time per lecture dropped from 11.0 hours (SD = not reported) with traditional methods to 4.2 hours (SD = not reported) when using AI assistance. The reduction was 62%, calculated as $(11.0 - 4.2) / 11.0 = 0.62$.

A paired t-test showed this difference was statistically significant ($p < .01$, with a large effect size expected based on Cohen's d).

Instructional quality

Mean quality scores (averaged across the five criteria in Table I) were:

Traditional preparation: $M = 3.81$ ($SD = 0.42$)

AI-assisted preparation: $M = 3.74$ ($SD = 0.48$)

The difference was not significant ($t(11) = 0.84$, $p = 0.42$, $d = 0.15$). So using AI did not lower the perceived quality of instruction.

Error analysis

We looked closely at the parts of AI-generated explanations that instructors had to correct. About 19.7% of those explanation units needed a manual fix. The errors fell into three main types, shown in Table II.

TABLE II. ERROR CATEGORIES

Error category	Frequency (%)	Example
Mathematical inaccuracies	38%	Incorrect rotation matrix order; inconsistent scaling formulas across steps
Faulty assumptions in transformations	34%	Assuming column-major coordinate system without specification; wrong transformation sequence (translation vs. scaling)
Incomplete / missing reasoning steps	28%	Omitting the need to convert between screen and world coordinates; stating "the object moves" without specifying relative vs. absolute

All errors were corrected by instructors before final delivery; none reached students.

Discussion

Our results show that generative AI tools, in this case ChatGPT-4 and GitHub Copilot, can make preparing teaching materials for 2D computer graphics much faster. A 62% drop in preparation time means AI is especially good at helping with structuring content, generating code, and making first-draft explanations.

Importantly, the speed gain did not come with a measurable drop in quality. That matches other recent work suggesting that generative AI works best as a cognitive support tool, not as a replacement for teaching expertise [3]. Looking through the lens of cognitive load theory [7], AI seems to take over some routine mental work, leaving instructors more time to focus on bigger-picture teaching decisions.

But there is also a clear limitation. Finding that about one in five AI outputs is wrong or incomplete shows that even with good prompts, the model is not yet consistent enough for specialized educational use. These mistakes showed up most often in mathematically tricky areas like geometric transformations and step-by-step algorithmic explanations.

This problem has been seen before in computing education research, which has reported that AI-generated code explanations vary in quality and can have conceptual mistakes that people have to catch [6]. In fields where being precise matters a lot, such inconsistencies can hurt both content accuracy and how well students understand.

Taken together, these findings suggest that fully automated lecture preparation is not yet possible for demanding technical subjects. Instead, a human-in-the-loop model works better: AI speeds up the creation of draft materials, but instructors stay responsible for checking, fixing, and making sure everything fits the learning goals.

In this hybrid model:

AI acts as a content accelerator, cutting routine workload.

Human instructors act as quality controllers, keeping things accurate and educationally sound.

This split of roles fits a broader trend in educational technology, moving from full automation toward augmentation, where AI is a helper, not an independent author.

Conclusion

This study provides real evidence that generative AI tools like ChatGPT-4 and GitHub Copilot can greatly improve how efficiently instructors prepare lectures for 2D computer graphics. Preparation time dropped by more than 60%, while the perceived quality of instruction stayed about the same.

AI seems to work mainly as an efficiency tool, automating routine parts of content creation such as structuring explanations, writing example code, and making first drafts. At the same time, we found a critical weakness: AI outputs

can have conceptual and mathematical errors that need a human to review.

So bringing AI into educational content development should not be seen as replacing instructor expertise, but rather as adding a complementary way to augment it. The teacher's role is still essential for making sure content is accurate, pedagogically coherent, and aligned with learning goals.

From a practical viewpoint, our results support using a hybrid human–AI workflow, where AI helps generate content faster and educators keep the final say over correctness and fit.

Limitations and future work

This study has several limits. The sample was very small (only two instructors), and the design was exploratory, so the findings should be taken as suggestive, not generalizable. Also, the work was done in just one course, which limits how far we can extend the conclusions.

Later studies could use more instructors from different backgrounds, cover a wider range of courses and subjects, follow students over a longer period to see how AI-assisted teaching affects their learning, and create a standard checklist for checking the quality of AI-generated educational content.

[1] Albadarin, Y., Saqr, M., Pope, N., & Tukiainen, M. (2024). A systematic review of ChatGPT in education. *Discover Education*.
[2] Lee, A. V. Y., Tan, S. C., & Teo, C. L. (2023). Generative AI for learning. *Smart Learning Environments*.
[3] Sharma, H., et al. (2024). Generative AI in teaching. *Education Sciences*.
[4] van den Berg, G., & du Plessis, E. (2023). AI in lesson planning. *Education Sciences*.
[5] Denny, P., et al. (2024). Generative AI in education. arXiv.
[6] Zastudil, C., et al. (2023). AI in computing education. arXiv.
[7] John Sweller. (1988). Cognitive Load Theory.
[8] Richard E. Mayer. (2009). Multimedia Learning.
[9] OpenAI. (2023). GPT-4 Technical Report.
[10] GitHub. (2023). Copilot Documentation.

DistilBERT Tabanlı Metin Sınıflandırmada LoRA ve Full Fine-Tuning Yaklaşımlarının Karşılaştırılması

Hülya Saygılı
Karayolları 8. Bölge Müdürlüğü
Elazığ, Türkiye
0000-0003-1926-9918

Cemile İnce
İnönü Üniversitesi Bilgi İşlem Daire Başkanlığı,
Malatya, Türkiye
0000-0002-4638-8501

Özet*—Bu çalışmada, metin sınıflandırma problemi için DistilBERT tabanlı iki farklı ince ayar yaklaşımı olan Full Fine-Tuning (Full FT) ve LoRA yöntemleri karşılaştırılmıştır. Deneyler, dört haber sınıfından oluşan AG News veri kümesi üzerinde gerçekleştirilmiştir. Rastgele seçilen 1000 örnek ile yapılan deneylerde; modeller doğruluk, kayıp, kesinlik, duyarlılık ve F1-skoru ölçütleri ile değerlendirilmiştir. Sonuçlar, Full FT modelinin eğitim sürecinde daha yüksek performans gösterdiğini; ancak test sonuçları açısından LoRA ile benzer düzeyde performansa sahip olduğunu göstermektedir. Ayrıca sınıf bazlı değerlendirmelerde Full FT modelinin daha dengeli sonuçlar ürettiği görülmüştür. Buna rağmen, LoRA yönteminin daha az parametre güncellemesine rağmen rekabetçi performans sergilemesi, parametre-verimli ince ayar yöntemlerinin metin sınıflandırma problemlerinde etkili biçimde kullanılabileceğini göstermektedir.

Anahtar Kelimeler: DistilBERT, LoRA, metin sınıflandırma, Transformer, doğal dil işleme.

***Abstract*—In this study, two different fine-tuning approaches based on DistilBERT, namely Full Fine-Tuning (Full FT) and LoRA, were compared for the text classification problem. The experiments were conducted on the AG News dataset, which consists of four news categories. Using 1,000 randomly selected samples, the models were evaluated based on accuracy, loss, precision, recall, and F1-score metrics.**
The results indicate that the Full FT model achieved higher performance during the training phase; however, it demonstrated a performance level similar to that of LoRA in terms of the test results. In addition, class-wise evaluations revealed that the Full FT model produced more balanced outcomes. Nevertheless, the fact that the LoRA method achieved competitive performance despite updating significantly fewer parameters suggests that parameter-efficient fine-tuning methods can be effectively utilized for text classification tasks.*

Keywords: DistilBERT, LoRA, text classification, Transformer, natural language processing

I. GIRIŞ

Doğal Dil İşleme (NLP) alanındaki araştırmalar, Transformer tabanlı mimariler [1] ve bu mimariler üzerine inşa edilen Büyük Dil Modellerinin (LLM) elde ettiği yüksek başarılar sayesinde önemli ölçüde hız kazanmıştır. BERT, RoBERTa ve DistilBERT gibi modeller, geniş ölçekli metin veri kümeleri üzerinde gerçekleştirilen ön eğitim (pre-training) süreçleri sayesinde dilin anlamsal ve bağlamsal yapısını öğrenmekte; ardından metin sınıflandırma, duygu analizi ve özetleme gibi belirli görevler için ince ayar yöntemiyle (Fine-tuning Strategies) özelleştirilmektedir [2,3]. Geleneksel tam ince ayar süreci, devasa parametre sayısının güncellenmesi nedeniyle hem ciddi bir hesaplama yükü oluşturmakta hem de veri kısıtı bulunan görevlerde modelin genelleme yeteneğini zayıflatarak aşırı uyuma yol açabilmektedir [4,5].

Gerçek hayatta, özellikle tıbbi raporlar, mühendislik dokümanları ya da düşük kaynaklı dillere ait veriler söz konusu olduğunda, büyük ve kapsamlı veri kümelerine ulaşmak çoğu zaman kolay olmamaktadır. Literatürde "küçük veri rejimi" (small-data regime) olarak adlandırılan bu durumlarda, milyonlarca parametreye sahip modeller eğitim verisini ezberleme eğilimi göstererek genelleme performansında belirgin düşüşler sergilemektedir [5]. Bu sorunu hafifletmek amacıyla geliştirilen Parametre Verimli İnce Ayar (Parameter-Efficient Fine-Tuning – PEFT) yöntemleri, modelin temel ağırlıklarını sabit tutarak yalnızca sınırlı sayıda ek parametrenin eğitilmesini sağlamaktadır. Bu yaklaşım hem hesaplama maliyetini düşürmekte hem de modelin genelleme kabiliyetini artırmaktadır [6].

Bu çalışmada, parametre verimliliği ile genelleme performansı arasındaki ilişki, AG News haber sınıflandırma veri seti üzerinde düşük veri koşulları altında incelenmiştir. Veri kümesi karıştırılarak, istatistiksel temsiliyeti korumak amacıyla yaklaşık eğitim veri kümesi içinden 1000 örneklik bir alt küme oluşturulmuştur. Çalışma kapsamında, DistilBERT mimarisi iki temel senaryo altında karşılaştırmalı olarak analiz edilmiştir:

- Modelin tüm parametrelerinin güncellendiği tam ince ayar (full fine-tuning)
- Düşük dereceli adaptasyon katmanlarının eğitildiği LoRA tabanlı yaklaşım.

LoRA tabanlı yaklaşımda, modelin tamamını güncellemek yerine yalnızca düşük dereceli adaptasyon katmanları eğitilmiştir. Bu yapılandırmada, dikkat (attention) katmanlarına düşük dereceli ek matrisler (low-rank, r=4) entegre edilmiş ve böylece modelin toplam parametrelerinin yalnızca yaklaşık %0,55'ine karşılık gelen kısmı (yaklaşık 368 bin parametre) güncellenmiştir. Bu durum, tam ince ayar (Full Fine-Tuning) yöntemine kıyasla önemli ölçüde hesaplama verimliliği sağlamıştır. Ayrıca eğitim süreci, Etiket Yumuşatma (Label Smoothing) ve Ağırlık Azaltımı (Weight Decay) gibi düzenlileştirme teknikleriyle desteklenerek aşırı uyum (overfitting) riski azaltılmıştır. Bu çalışma ile düşük veri koşullarında farklı ince ayar stratejilerinin genelleme

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

performansı üzerindeki etkilerinin karşılaştırmalı olarak ortaya konulması amaçlanmaktadır.

Çalışmanın geri kalan kısmı şu şekilde yapılandırılmıştır: İkinci bölümde ilgili literatür ve benzer çalışmalar incelenmekte; üçüncü bölümde önerilen yöntem, veri seti hazırlığı ve deneysel düzenek ayrıntılarıyla sunulmaktadır. Dördüncü bölümde elde edilen deneysel bulgular analiz edilerek tartışılmakta; son bölümde ise çalışma özetlenerek gelecek araştırmalar için potansiyel yönelimler sunulmaktadır.

II. ILIŞKILI ÇALIŞMALAR

Transformer mimarisi öncesinde dikkat mekanizması, genellikle yinelemeli sinir ağlarının (RNN) uzun vadeli bağımlılıkları çözmesine yardımcı olan ikincil bir bileşen olarak konumlandırılmıştır [7,8]. Buna karşın Vaswani vd. (2017), dikkat mekanizmasını modelin merkezine yerleştirerek, yineleme ve evrişim işlemlerine duyulan ihtiyacı tamamen ortadan kaldıran, paralel işlemeye uygun bir mimari önermiştir [1].

Transformer tabanlı modellerin başarısına rağmen, bu yapıların yüksek parametre sayısı ve hesaplama maliyeti önemli bir sınırlılık oluşturmaktadır. Bu soruna yönelik olarak Sanh vd. (2019), bilgi damıtma temelli olarak geliştirdikleri DistilBERT modeli ile BERT'in daha hafif ve hızlı bir sürümünü önermiştir. Çalışmada, model boyutunun yaklaşık %40 oranında küçültülmesine rağmen dil anlama performansının büyük ölçüde korunduğu ve çıkarım süresinin önemli ölçüde iyileştiği belirtilmiştir [3].

Benzer şekilde, Hu vd. (2021), büyük ölçekli önceden eğitilmiş dil modellerinin daha verimli biçimde uyarlanabilmesi amacıyla LoRA yöntemini önermiştir. Bu yöntemde, modelin önceden öğrenilmiş ağırlıkları sabit tutulmakta ve yalnızca Transformer katmanlarına eklenen düşük-rank matrisler eğitilmektedir. Çalışmada, bu yaklaşımın eğitilebilir parametre sayısını ve hesaplama maliyetini önemli ölçüde azalttığı, buna rağmen model performansını büyük ölçüde koruduğu ve bazı durumlarda daha başarılı sonuçlar üretebildiği gösterilmiştir [4]. LoRA'nın başarısını takiben, bellek verimliliğini 4-bit kuantizasyon ile maksimize eden QLoRA [9] ve parametre bütçesini katmanlara dinamik dağıtan AdaLoRA [10] gibi daha güncel yaklaşımlar geliştirilmiştir.

Büyük dil modellerinin eğitim ve uyarlama süreçlerinde kullanılan düzenlileştirme teknikleri de performans üzerinde önemli etkiye sahiptir. Bu bağlamda Müller vd. (2019), çok sınıflı sinir ağlarında kullanılan etiket düzeltme (label smoothing) yönteminin yalnızca genelleme performansını artırmakla kalmadığını, aynı zamanda model kalibrasyonunu da iyileştirdiğini göstermiştir [11].

Liu vd. (2022), yaptıkları çalışmada az örnekli bağlam içi öğrenme (few-shot in-context learning, ICL) ile parametre-verimli ince ayar (PEFT) yöntemlerini karşılaştırarak PEFT yaklaşımlarının hem doğruluk hem de hesaplama maliyeti açısından daha avantajlı olduğunu ortaya koymuştur. Aynı çalışmada, çok az sayıda ek parametre ile etkili sonuçlar üreten $(IA)^3$ yöntemi ile görev-özel ayar gerektirmeden yeni görevlere uygulanabilen T-Few yaklaşımı önerilmişlerdir [6].

III. MATERYAL VE YÖNTEM

A. Veri Seti (AG News)

AG News, Xiang Zhang tarafından oluşturulmuş, AG haber arşivinden1 milyondan fazla haber metni içeren bir metin sınıflandırma veri kümesidir. Veri kümesi, özgün arşivdeki dört haber kategorisinin (Dünya, Spor, İş ve Bilim/Teknoloji) seçilmesiyle hazırlanmıştır. Her bir sınıf 30.000 eğitim ve 1.900 test örneği içermekte olup, toplamda 120.000 eğitim ve 7.600 test örneğinden oluşmaktadır [12].

B. Model Mimarisi

DistilBERT, bilgi damıtma (knowledge distillation) yaklaşımıyla önceden eğitilmiş, daha küçük boyutlu ve daha hızlı çalışan bir dil modelidir. Ön eğitim aşamasında kullanılan üçlü kayıp fonksiyonu sayesinde, daha düşük hesaplama maliyetiyle büyük Transformer tabanlı modellere benzer bir performans sunabilmektedir [13].

Bu çalışmada, metin sınıflandırma problemi için DistilBERT tabanlı bir Transformer mimarisi kullanılmıştır. Temel model olarak Hugging Face kütüphanesinde yer alan distilbert-base-uncased önceden eğitilmiş dil modeli tercih edilmiştir. Model, giriş metinlerini tokenize edilmiş biçimde işleyerek bağlamsal temsil vektörleri üretmekte ve son katmanda yer alan sınıflandırma başlığı aracılığıyla dört haber sınıfından birine atama yapmaktadır.

Model girişinde her metin örneği en fazla 128 token ile sınırlandırılmış ve gerekli durumlarda dinamik doldurma (dynamic padding) uygulanmıştır. Sınıflandırma görevi için modelin çıkış katmanı 4 sınıflı softmax yapısına göre düzenlenmiştir. Bu çalışma, aynı temel mimari üzerinde iki farklı ince ayar yaklaşımı olan Full Fine-Tuning ve LoRA ile karşılaştırmıştır.

C. Full Fine-Tuning (Full FT)

Bu yaklaşımda, DistilBERT modelinin tüm parametreleri eğitime açılmıştır. Başka bir ifadeyle, önceden eğitilmiş Transformer katmanları ile sınıflandırma katmanı birlikte güncellenmiştir. Bu yöntem, modelin hedef göreve ait veri kümesine daha iyi uyum sağlamasına imkan tanımakla birlikte, daha yüksek hesaplama maliyeti ve daha fazla sayıda eğitilebilir parametre gerektirmektedir.

D. LoRA Tabanlı İnce Ayar

Low-Rank Adaptation (LoRA) yönteminin kullanıldığı ikinci yaklaşımda, önceden eğitilmiş DistilBERT model ağırlıkları büyük ölçüde sabit tutulmuş; yalnızca dikkat mekanizmasındaki belirli doğrusal projeksiyon katmanlarına düşük dereceli güncelleme matrisleri eklenmiştir. Bu çalışmada LoRA, self-attention yapısında sorgu ve değer vektörlerinin oluşturulmasında kullanılan katmanlara uygulanmıştır. LoRA yapılandırmasında rank değeri ($r=4$), ölçekleme katsayısı ($\alpha=16$) ve dropout oranı 0.3 olarak belirlenmiştir. Bu sayede modelin tüm parametrelerini güncellemek yerine yalnızca sınırlı sayıda ek parametre eğitilmiş, böylece hesaplama maliyeti azaltılırken görev performansı korunmuştur

E. Deneysel Ayarlar ve Hiperparametreler

Deneyler AG News eğitim veri kümesinden rastgele seçilen 1000 örnek üzerinde gerçekleştirilmiştir. Veri kümesi %70 eğitim, %10 doğrulama ve %20 test olacak şekilde ayrılmıştır. Giriş metinleri için maksimum dizi uzunluğu 128 token olarak belirlenmiş ve dinamik doldurma uygulanmıştır.

Çalışmada, DistilBERT tabanlı iki farklı ince ayar yaklaşımı karşılaştırılmıştır: Full Fine-Tuning (Full FT) ve LoRA. Her iki model 7 iterasyon (epoch) boyunca eğitilmiş, eğitim ve değerlendirme için batch size = 16 kullanılmıştır. Ayrıca weight decay = 0.05, warmup step = 50 ve label smoothing = 0.1 olarak ayarlanmıştır.

Öğrenme oranı olarak [1e-5, 2e-5, 3e-5, 1e-4, 2e-4, 3e-4] denenmiş ve en yüksek doğrulama başarımı veren değerler seçilmiştir. Model performansları doğruluk (accuracy) metriği ile değerlendirilmiş; ayrıca karmaşıklık matrisi (confusion matrix), sınıflandırma raporu (classification report), eğitim kaybı ve test performans eğrileri incelenmiştir.

IV. Bulgular

Tablo 1'de, LoRA ve Full FT modelleri test kümesinde aynı doğruluk değerine (%89.00) ulaşmıştır. Full FT modeli eğitim kümesinde daha yüksek doğruluk elde etmiş olsa da test performanslarının eşit olması LoRA yönteminin daha az parametre güncellemesiyle benzer genelleme başarısı sağlayabildiğini göstermektedir.

TABLE I. LoRA ve Full FT modellerinin performanslarının karşılaştırılması

Model	*Eğitim Kayıp*	*Test Kayıp*	*Doğrulama Kayıp*	*Eğitim Doğruluk %*	*Test Doğruluk %*	*Doğrulama Doğruluk %*
LoRA	0.1855	0.3108	0.3579	93.87	89.00	86.00
Full FT	0.1201	0.3070	0.3698	97.62	89.00	86.00

Tablo 2'de, Full FT modeli sınıf bazlı performans açısından daha dengeli sonuçlar üretirken, LoRA modeli de bazı sınıflarda benzer ve rekabetçi sonuçlar sergilemiştir. LoRA yöntemi, daha az parametre güncellemesine rağmen rekabetçi bir performans sergilediği görülmektedir.

TABLE II. LoRA ve Full FT modellerinin sınıf bazlı test performanslarının karşılaştırılması

Model	*Sınıf*	*Kesinlik*	*Duyarlılık*	*F1-Skoru*	*Destek*
LoRA	Dünya	0.9268	0.8261	0.8736	46
	Spor	0.9787	0.9200	0.9485	50
	İş	0.8000	1.0000	0.8889	44
	Bilim/Teknoloji	0.9123	0.8667	0.8889	60
Full FT	Dünya	0.8696	0.8696	0.8696	46
	Spor	1.0000	0.9200	0.9583	50
	İş	0.8889	0.9091	0.8989	44
	Bilim/Teknoloji	0.9206	0.9667	0.9431	60

Şekil 1 ve Şekil 2 incelendiğinde, Full FT modelinin eğitim sürecinde daha hızlı öğrenme gösterdiği; ancak test performansı açısından LoRA ile benzer sonuçlar verdiği görülmektedir.

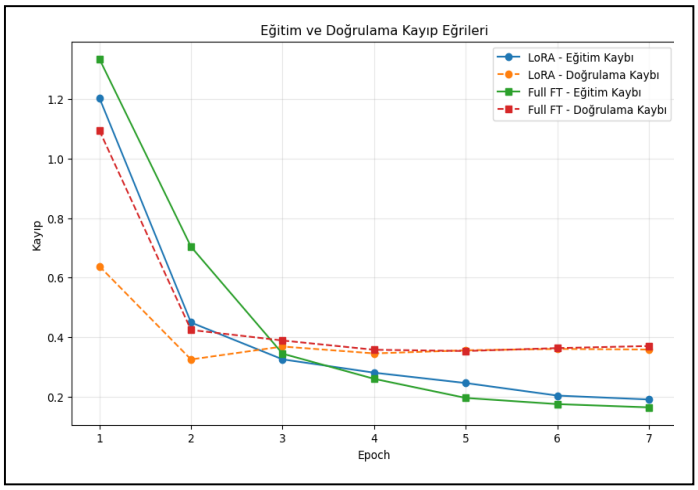


Fig. 1. LoRA ve FFT modellerine ait doğruluk eğrileri

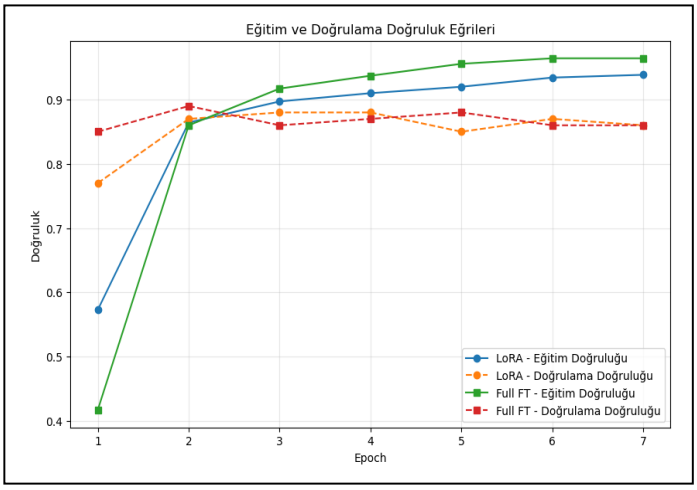


Fig. 2. LoRA ve Full FT modellerine ait eğitim ve test doğruluk eğrileri

Şekil 3 ve Şekil 4'teki karmaşıklık matrisleri incelendiğinde, her iki modelin de genel olarak başarılı sınıflandırma yaptığı görülmektedir. Full FT modeli sınıf ayrımında daha dengeli sonuçlar üretirken, LoRA modelinde özellikle İş ve Bilim/Teknoloji sınıfları arasında daha fazla karışıklık gözlenmiştir. Bu karışıklığın nedeninin bu iki kategorinin paylaştığı sektörel terminoloji ve teknik kelime dağarcığındaki yüksek benzerlikten ya da düşük rank değerinin (r=4) bu sınıflar arasındaki ince semantik ayrımları yakalamada kısıtlı kalmasından kaynaklanmış olabileceği düşünülmektedir.

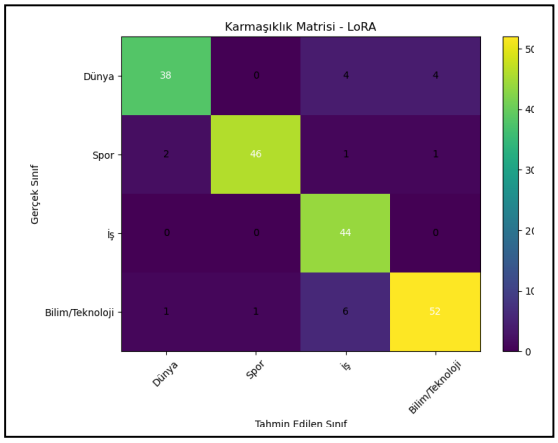


Fig. 3. LoRA karmaşıklık matrisi

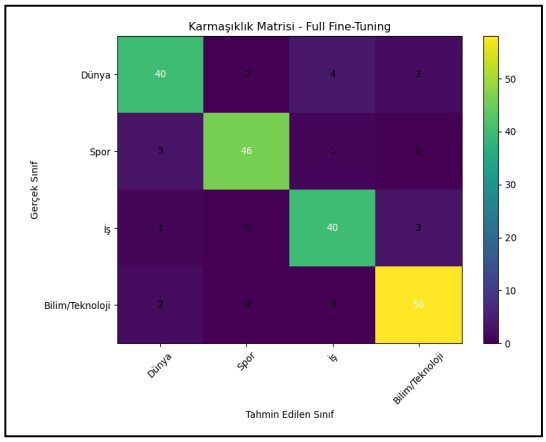


Fig. 4. Full FT karmaşıklık matrisi

V. SONUÇLAR

DistilBERT tabanlı metin sınıflandırma performansının incelendiği bu çalışmada, düşük veri rejiminde LoRA ve Full FT yaklaşımları analiz edilmiştir. Elde edilen bulgular, her iki modelin de %89.00 test doğruluğu ile benzer genelleme kabiliyeti sergilediğini ortaya koymuştur. Ancak Full FT modelinin eğitim kümesinde daha düşük kayıp (0.1201) ve daha yüksek doğruluk (%97.62) değerlerine ulaşması, tam ince ayar sürecinin sınırlı veri setlerinde aşırı uyum eğiliminin daha yüksek olduğunu, LoRA'nın ise parametre kısıtı sayesinde daha dirençli bir öğrenme sergilediğini göstermektedir.

Sınıf bazlı analizlerde Full FT modelinin özellikle "İş" ve "Bilim/Teknoloji" kategorileri arasında daha dengeli bir ayrım yapabildiği gözlenmiştir. LoRA modelinde bu iki sınıf arasındaki karışıklığın, seçilen düşük rank değerinden (r=4) kaynaklanabileceği değerlendirilmektedir. LoRA'nın, düşük kaynaklı NLP uygulamaları açısından güçlü bir alternatif olarak öne çıkmaktadır.

Bu çalışma, DistilBERT gibi zaten optimize edilmiş modellerde bile LoRA'nın %99,45 oranında parametre tasarrufu sağlayarak performansı koruyabildiğini göstermektedir. Bu bulgu, özellikle kenar cihazlarda (edge devices) çalışan gerçek zamanlı NLP uygulamaları için önemli bir enerji ve bellek verimliliği potansiyeline işaret etmektedir.

Gelecek çalışmalarda, rank değerinin (r) duyarlılık analizinin yapılması ve modelin tıbbi metinler gibi daha özel alanlarda test edilmesi planlanmaktadır.

KAYNAKÇA

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems, 2017.

[2] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of the North American Chapter of the Association for Computational Linguistics, 2019.

[3] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," 2019, arXiv:1910.01108. [Online]. Available: https://arxiv.org/abs/1910.01108

[4] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," 2021, arXiv:2106.09685. [Online]. Available: https://arxiv.org/abs/2106.09685

[5] L. Brigato and L. Iocchi, "A close look at deep learning with small data," in International Conference on Pattern Recognition (ICPR), 2021.

[6] H. Liu, D. Tam, M. Muqeeth, J. Mohta, T. Huang, M. Bansal, and C. Raffel, "Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning," 2022, arXiv:2205.05638. [Online]. Available: https://arxiv.org/abs/2205.05638

[7] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, arXiv:1409.0473. [Online]. Available: https://arxiv.org/abs/1409.0473

[8] T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," 2015, arXiv:1508.04025. [Online]. Available: https://arxiv.org/abs/1508.04025

[9] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, "QLoRA: Efficient finetuning of quantized LLMs," 2023, arXiv:2305.14314. [Online]. Available: https://arxiv.org/abs/2305.14314

[10] Q. Zhang, M. Chen, A. Bukharin, P. He, Y. Cheng, W. Chen, and T. Zhao, "Adaptive budget allocation for parameter-efficient fine-tuning," 2023, arXiv:2303.10512. [Online]. Available: https://arxiv.org/abs/2303.10512

[11] R. Müller, S. Kornblith, and G. E. Hinton, "When does label smoothing help?" 2019, arXiv:1906.02629. [Online]. Available: https://arxiv.org/abs/1906.02629

[12] Zhang, X., Zhao, J.J., & LeCun, Y. (2015). Character-level Convolutional X. Zhang, J. J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in Advances in Neural Information Processing Systems, 2015.

[13] Hugging Face, "DistilBERT documentation," [Online]. Available: https://huggingface.co/docs/transformers/model_doc/distilbert. Accessed: Mar. 25, 2026.

Dijital Çağda Kültürel Diplomasi: Yapay Zekâ Temelli İletişim Stratejileri

Tolkun Bekturgan kızı
Kırgızistan-Türkiye Manas Üniversitesi
İletişim Fakültesi / Gazetecilik Bölümü
Bişkek / Kırgızistan
tolkun.bekturgan@manas.edu.kg /
ORCID: 0000-0002-0596-4090

***Öz*- Bu bildiri, dijital çağın getirdiği köklü dönüşümler çerçevesinde kültürel diplomasinin nasıl yeniden biçimlendiği ve yapay zekâ (YZ) temelli iletişim stratejilerinin bu dönüşüme nasıl katkı sağladığını teorik ve betimsel bir perspektiften ele almaktadır. Geleneksel kültürel diplomasi anlayışının sınırlarını ve araçlarını sorgulamakla başlayan çalışma, dijital platformların devlet ve devlet dışı aktörlere sunduğu olanaklarla birlikte YZ uygulamalarının bu alanda yarattığı paradigmatik kaymaları incelemektedir.**

Nitel betimsel analiz yöntemiyle yürütülen bu çalışmada; çeşitli ulusal ve ulusüstü aktörlerin dijital kültür diplomasisi pratikleri mercek altına alınmış, özellikle büyük dil modelleri, algoritmik içerik önerileri ve kişiselleştirilmiş iletişim stratejilerinin kamusal diplomasi üzerindeki etkileri tartışılmıştır. Bulgular, yapay zekânın kültürel diplomasiyi hem demokratikleştirici hem de manipülatif bir araç olarak işlev gördüğüne işaret etmektedir. Bu çelişkili tablo, etik, egemenlik ve kültürel kimlik ekseninde kapsamlı bir değerlendirmeyi zorunlu kılmaktadır.

Sonuç itibarıyla bildiri; devletlerin, uluslararası örgütlerin ve sivil toplum kuruluşlarının dijital kültürel diplomasiyi etkin, sorumlu ve şeffaf bir biçimde yürütebilmeleri için geliştirmeleri gereken kurumsal ve stratejik çerçeveye dair öneriler sunmaktadır.

Anahtar Kelimeler— Kültürel diplomasi, yapay zekâ, dijital iletişim, yumuşak güç, uluslararası ilişkiler

I. GIRIŞ

Uluslararası ilişkiler tarihinde diplomasi, yüzyıllar boyunca devletlerin birbirleriyle etkileşim kurduğu ve ortak çıkarların müzakere edildiği temel mekanizma olarak konumlanmıştır. Ancak 21. yüzyılın başından itibaren yaşanan dijital dönüşüm, bu köklü mekanizmayı kökten sarsmıştır. Bilgi ve iletişim teknolojilerinin evrensel ölçekte yaygınlaşması, bireylerin, sivil toplum örgütlerinin ve ticari aktörlerin diplomatik sürece katılımını kolaylaştırmış; böylece diplomasi, yalnızca devlet aktörlerine özgü bir faaliyet olmaktan çıkmıştır.

Bu dönüşümün en çarpıcı boyutlarından birini kültürel diplomasi alanında gözlemlemek mümkündür. Joseph Nye'ın [1] 'yumuşak güç' kavramı çerçevesinde değerlendirilen kültürel diplomasi; bir ülkenin kültürünü, değerlerini ve kurumlarını başka toplumlarla paylaşarak uluslararası kamuoyunda olumlu bir imaj oluşturma ve etki alanını genişletme sürecini ifade etmektedir. Geleneksel anlamda bu süreç; kültür merkezleri, değişim programları, uluslararası yayın kuruluşları ve sanat etkinlikleri gibi araçlar aracılığıyla yürütülmekteydi.

Ne var ki dijital çağ, söz konusu araçları hem çeşitlendirmiş hem de kökten dönüştürmüştür. Sosyal medya platformları, akış hizmetleri, dijital oyunlar ve sanal gerçeklik uygulamaları, kültürel içeriklerin küresel ölçekte anında yayılmasına zemin hazırlamıştır. Bu sürecin en kritik boyutunu ise yapay zekâ teknolojilerinin devreye girişi oluşturmaktadır. Doğal dil işleme, makine öğrenimi ve büyük veri analitiğine dayanan YZ sistemleri; hedef kitleyi tanımlamaktan kültürel içerik üretimine, gerçek zamanlı çeviri hizmetlerinden kişiselleştirilmiş iletişim stratejilerine uzanan geniş bir yelpazede diplomatik pratiği yeniden şekillendirmektedir.

Bu bildiri, söz konusu dönüşümü teorik ve betimsel bir perspektiften ele alarak yapay zekânın kültürel diplomaside yarattığı paradigma kaymasını analiz etmeyi amaçlamaktadır. Çalışmanın temel sorusu şudur: Yapay zekâ temelli iletişim stratejileri, geleneksel kültürel diplomasi anlayışını nasıl dönüştürmekte ve bu dönüşüm uluslararası ilişkiler açısından ne gibi fırsatlar ve tehditler barındırmaktadır?

II. KURAMSAL ÇERÇEVE

A. Kültürel Diplomasi: Kavramsal Bir Değerlendirme

Kültürel diplomasi kavramı, akademik literatürde çeşitli şekillerde tanımlanmıştır. Milton Cummings [2], kültürel diplomasiyi 'farklı toplumlar arasında karşılıklı anlayışı geliştirmek amacıyla fikirlerin, bilginin, sanatın ve kültürün diğer unsurlarının değiş tokuşu' olarak tanımlamaktadır. Bu perspektiften hareketle kültürel diplomasiyi, salt siyasi ya da ekonomik hesapların ötesinde kültürel bağ ve sempati inşa etmeyi hedefleyen bir yumuşak güç aracı olarak okumak mümkündür.

Geleneksel kültürel diplomasi modelleri, büyük ölçüde devlet merkezli ve tek yönlü bir iletişim paradigmasına dayanmaktadır. Bu modelde ulusal kültürel ajanslar, büyükelçilikler ve hükümet destekli vakıflar gibi kurumlar, belirli kültürel mesajları hedef ülkelere ulaştırma işlevini üstlenmekteydi. Fransa'nın Alliance Française, Almanya'nın Goethe-Institut veya Türkiye'nin Yunus Emre Enstitüsü, bu geleneksel yaklaşımın tipik örnekleri arasında sayılabilir.

Ancak Melissen [3], yeni kamu diplomasisi tartışmasında bu tek yönlü modelin sınırlarına dikkat çekmektedir. Ona göre etkili kültürel diplomasi, diyaloğu, ortak anlam üretimini ve dinleyici kitlelerin aktif katılımını gerektirmektedir. Dijital teknolojiler ve yapay zekâ, tam da bu anlamda daha diyalogcu ve katılımcı bir kültürel diplomasi pratiğinin zeminini hazırlamaktadır.

B. Dijital Dönüşüm ve Uluslararası İletişim

Dijital dönüşüm, uluslararası iletişimin yapısını köklü biçimde değiştirmiştir. Castells'in [4] 'ağ toplumu' kuramı, bu dönüşümü kavramsallaştırmak için sıklıkla başvurulan bir çerçeve sunmaktadır. Castells'e göre bilgi ağları, güç ve

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

etkinin yeni dağılım biçimlerini belirlemekte; devletlerin iletişim süreçleri üzerindeki geleneksel tekeli aşınmaktadır.

Twitter (X), Facebook, Instagram, TikTok ve YouTube gibi sosyal medya platformları, hükümetlere ve uluslararası kuruluşlara doğrudan kitlelerle iletişim kurabilme imkânı tanımaktadır. Bu durum, geleneksel diplomatik kanalları devre dışı bırakmasa da tamamlamakta ve kimi zaman rakip mesajların ortaya çıkmasına yol açmaktadır. Pamment [5], 'dijital kamu diplomasisi' kavramıyla bu yeni ortamda etkili bir varlık sürdürebilmek için stratejik bir dijital iletişim yönetimi anlayışının zorunlu hale geldiğini vurgulamaktadır.

Öte yandan dijital alan, kültürel diplomasiye özgü yeni gerilimler de üretmektedir. Nitekim 'filtre balonları' ve 'yankı odaları', hedef kitlenin karşıt kültürel mesajlara maruz kalma ihtimalini daraltmaktadır. Bunun yanı sıra dezenformasyon ve yanlış bilginin yayılması, kültürel diplomasi girişimlerine yönelik güveni zedeleyebilmekte; algoritmaların kültürel içerikler üzerindeki belirleyici etkisi ise bu alanda neyin görünür olduğunu denetim altında tutma kapasitesini devletlerin elinden almaktadır.

C. Yapay Zekânın İletişim Stratejilerindeki Rolü

Yapay zekâ teknolojileri, son on yılda iletişim stratejileri açısından devrimsel bir etki yaratmıştır. Doğal dil işleme (NLP), bilgisayarlı görü ve pekiştirmeli öğrenme başta olmak üzere çeşitli YZ alt disiplinleri, birden fazla alanda iletişimin dönüşümünü hızlandırmaktadır.

Büyük dil modelleri (BDM), kültürel diplomatik alanda özellikle dikkat çekici uygulamalara zemin hazırlamaktadır. GPT-4, Claude, Gemini ve benzeri modeller; hedef dillerde ve kültürel bağlamlarda nüanslı metinler oluşturabilmekte, karmaşık kültürel referansları yorumlayabilmekte, diplomatik iletişimi gerçek zamanlı olarak çevirebilmekte ve farklı izlerkitlelere yönelik kişiselleştirilmiş kültürel anlatılar geliştirebilmektedir.

Duygu analizi ve kitle araştırması bağlamında YZ sistemleri, büyük veri kümelerini işleyerek hedef kitleler hakkında kültürel tercihler, değerler ve duygu durumları gibi alanlarda kapsamlı içgörüler üretebilmektedir. Bu tür bilgiler; kültürel mesajların belirlenen izlerkit lere göre uyarlanmasında, kültürel diplomatik kampanyaların sonuçlarının izlenmesinde ve kültürel çatışma ile gerilim alanlarının tespit edilmesinde kritik bir işlev görmektedir.

III. YAPAY ZEKÂ TEMELLI KÜLTÜREL DIPLOMASI PRATIKLERI

A. Kişiselleştirilmiş Kültürel İletişim

Geleneksel kültürel diplomasinin belki de en köklü sınırı, kitlesel ve homojen bir iletişim anlayışına dayanmasıydı. YZ, bu anlayışı kökten dönüştürerek büyük ölçekli kişiselleştirmeye olanak tanımaktadır. Bu bağlamda algoritmik içerik önerileri belirleyici bir işlev üstlenmektedir: Netflix, Spotify ve YouTube gibi platformların kullanıcı tercihlerine dayalı içerik öneri sistemleri, kültürel içeriklerin tüketilme biçimini doğrudan şekillendirmektedir. Devletler ve kültürel kurumlar, kendi kültürel içeriklerini bu platformlara yerleştirerek algoritmik dağıtım altyapısından etkin biçimde yararlanabilmektedir.

Öte yandan hedefli dijital kampanyalar açısından YZ destekli reklamcılık araçları, demografik veriler, coğrafi konum bilgisi ve ilgi alanı profilleri temelinde kültürel mesajların son derece hassas biçimde hedeflenmesine imkân tanımaktadır. Ulusal kültürel ajanslar bu araçları; sanatsal etkinliklerin tanıtımında, kültürel değişim programlarının duyurulmasında ve turistik kampanyaların yürütülmesinde giderek artan bir sıklıkla kullanmaktadır. Güney Kore Kültür ve Turizm Bakanlığı'nın K-Pop ile K-drama'ya yönelik YZ destekli küresel pazarlama kampanyaları, bu eğilimin somut bir örneğini oluşturmaktadır.

Bununla birlikte kişiselleştirilmiş kültürel iletişimin önemli etik sorular barındırdığı da göz ardı edilemez. Hedefleme algoritmaları, bireylerin mevcut kültürel önyargılarını pekiştirerek merak duygularını ve farklı kültürel deneyimlere açıklıklarını köreltiyor olabilir. Ayrıca kişisel verilerin kültürel diplomatik amaçlarla kullanılması, mahremiyet ve rıza konusunda ciddi soru işaretleri doğurmaktadır.

B. Yapay Zekâ Destekli Kültürel Çeviri ve Erişilebilirlik

Dil, kültürel diplomasinin önündeki en kadim engellerden birini oluşturmaktadır. YZ destekli çeviri teknolojileri bu engeli kısmen aşmakla birlikte, beraberinde yeni zorlukları da gündeme getirmektedir. Otomatik çeviri hizmetleri bağlamında DeepL, Google Translate ve Microsoft Translator gibi YZ tabanlı araçlar, kültürel içeriklerin çok dilli yayılımını kolaylaştırmaktadır. Bu hizmetler, kültürel kuruluşların düşük maliyetle küresel erişim sağlamasına zemin hazırlarken bazı eleştirmenler, çevirilerin kültürel incelikleri, deyimleri ve mizahi unsurları her zaman yeterince aktaramadığını öne sürmektedir.

Gerçek zamanlı çeviri kapasitesi ise özellikle canlı kültürel etkinlikler ve sanal diplomasi toplantıları açısından kritik bir öneme sahiptir. YZ destekli simultane çeviri, daha fazla aktörün uluslararası kültürel diyaloglara katılımını mümkün kılmaktadır. Bununla birlikte mevcut en gelişmiş sistemler bile kültürel bağlamı ve konuşma dilinin nüanslarını tam olarak kavramakta yetersiz kalmaya devam etmektedir. Kültürel adaptasyon konusunda ise en ileri YZ sistemleri bile içeriklerin yalnızca dilsel değil, kültürel açıdan da uyarlanmasında sınırlı bir başarı sergilemektedir. Bir ülkede yankı uyandıran mizah, değerler veya anlatı biçimleri, başka bir kültürel bağlamda beklenen etkiyi yaratmayabilir.

C. YZ Üretimli Kültürel İçerik

YZ'nin kültürel diplomasi açısından belki de en tartışmalı boyutunu, kültürel içerik üretimindeki rolü oluşturmaktadır. Sanat ve yaratıcı üretim alanında YZ araçları artık görsel sanat, müzik, şiir ve edebi içerik üretebilmektedir. Bu durum hem fırsatlar hem de köklü sorular barındırmaktadır: YZ tarafından üretilen kültürel içerik, gerçek bir kültürel ifade sayılabilir mi? Bu içerikler, kültürel diplomatik mesajları taşıma işlevini etkili biçimde yerine getirebilir mi? Kamusal diplomasi anlatıları konusunda ise bazı hükümetler, ulusal kimliklerini ve değerlerini yabancı kitlelere aktarmak amacıyla YZ'yi kamusal diplomasi anlatıları üretmede kullanmaktadır. Bu yaklaşım ölçeklenebilirlik ve maliyet etkinliği sunsa da özgünlük ve manipülasyon kaygılarını beraberinde getirmektedir.

IV. Etiik ve Egemenlik Boyutları

A. Kültürel Emperyalizm ve Teknolojik Hegemonya

Yapay zekâ temelli kültürel diplomasinin en kritik eleştirilerinden biri, mevcut güç asimetrilerini pekiştirme potansiyeliyle ilgilidir. YZ geliştirme kapasitesi, büyük ölçüde Amerika Birleşik Devletleri ve Çin gibi teknoloji alanında öncü konumdaki birkaç ülkede yoğunlaşmış durumdadır. Bu asimetrik yapı, söz konusu ülkelerin algoritmik sistemler aracılığıyla küresel kültürel iletişim üzerinde orantısız bir etki gücü elde etmelerine zemin hazırlamaktadır.

Schiller'in [6], medya emperyalizmi tezini günümüz bağlamında yeniden okuyacak olursak, YZ aracılı kültürel nüfuz yeni bir teknolojik emperyalizm biçimini temsil etmekte olabilir. Global platformlar; içerikleri sınıflandıran, görünürlük hiyerarşileri oluşturan ve kültürel mesajları şekillendiren algoritmalar geliştirmekte ve bu süreçte belirli kültürel perspektifler lehine ya da aleyhine sonuçlar doğurabilmektedir.

Tomlinson [7], kültürel küreselleşme sürecinde kültürel homojenleşme ile çeşitlilik arasındaki gerilime dikkat çekmektedir. YZ'nin kültürel diplomaside sahneye çıkışı, bu gerilimi daha da derinleştirebilir: Daha küçük ve azınlık kültürlerine ait içerikler, algoritmik görünürlük yarışında hâkim kültürlerin gölgesinde kalma riskiyle yüz yüzedir.

B. Dezenformasyon ve Manipülasyon Riskleri

Yapay zekânın kültürel diplomatik araç olarak kullanımı, dezenformasyon üretme kapasitesiyle doğrudan örtüşmektedir. Deepfake teknolojisi, kültürel kişiliklere ait sahte video ve ses içerikleri üretilmesine olanak tanıyarak kültürel diplomasinin güvenilirliğini ciddi ölçüde zedeleyebilir. Nitekim kültürel temsilcilerin ya da ulusal liderlerin manipüle edilmiş görüntüleri, uluslararası ilişkilerde derin bir güvensizlik ortamı yaratma potansiyeli taşımaktadır.

Koordineli YZ destekli dezenformasyon kampanyaları, gerçek kültürel etkileşimi taklit ederek kamuoyunu ve kültürel algıları manipüle edebilmektedir. Bu tür operasyonlar, uluslararası toplumun kültürel diplomasi girişimlerinin özgünlüğünü sorgulamasına yol açmaktadır. Benzer biçimde, sanat eserlerini veya kültürel üretimi YZ ile hızla kopyalamak mümkün olduğundan, fikri mülkiyet hakları ve kültürel miras konusundaki anlaşmazlıklar da kritik bir etik meseleye dönüşmektedir.

C. Veri Egemenliği ve Kültürel Mahremiyet

Kültürel tercihleri ve davranışları kapsayan kişisel verilerin toplanması, diplomatik amaçlarla kullanıldığında ciddi mahremiyet riskleri doğurmaktadır. Hedef ülkelerdeki bireyler hakkında kapsamlı veri tabanları oluşturulması, bu verilerin yalnızca kültürel erişimi kolaylaştırmak için değil, hassas sosyal ve siyasi manipülasyon amacıyla da kullanılabileceğine işaret etmektedir.

Zuboff'un [8] 'gözetim kapitalizmi' kavramı, kültürel diplomasi bağlamında yeniden çerçevelendirildiğinde daha da çarpıcı bir anlam kazanmaktadır. Kültürel tercihlerin ve tüketim alışkanlıklarının sistematik biçimde toplanması ve analiz edilmesi, kültürel alanda gözetim temelli bir iktidar modeli olan 'kültürel gözetim' sorununu gündeme getirmektedir.

V. Güncel Uygulamalar: Nitel Betimsel Analiz Kaynakça

A. Çin'in Dijital Kültürel Diplomasisi

Çin Halk Cumhuriyeti, dijital platformlar ve YZ aracılığıyla kültürel diplomasi kapasitesini sistematik biçimde geliştirmektedir. Bu kapsamda 'yeni dönem kamu diplomasisi' stratejisi; Çin kültürü, dili ve değerlerinin küresel ölçekte yayılmasını destekleyen kapsamlı dijital altyapı yatırımlarını içermektedir.

TikTok'un ana şirketi ByteDance ve WeChat'ın sahibi Tencent başta olmak üzere Çin merkezli dijital platformlar, küresel kültürel içerik dağıtımında belirleyici konumlara ulaşmıştır. Bu durum, söz konusu platformların algoritmalarının hangi kültürel içerikleri öne çıkardığı ve bunun uluslararası kültürel algılar üzerindeki olası etkileri konusunda ciddi sorular doğurmaktadır.

Batılı hükümetlerin bu platformlara ilişkin endişeleri, YZ temelli kültürel diplomasinin güvenlik boyutlarını giderek daha fazla gündeme taşımaktadır. Nitekim ABD, Avustralya ve Hindistan gibi ülkelerdeki yasaklar veya kısıtlamalar, dijital kültürel alandaki jeopolitik gerilimlerin somut yansımalarıdır.

B. Güney Kore'nin Hallyu Stratejisi ve YZ Entegrasyonu

Güney Kore'nin 'Kore Dalgası' (Hallyu) fenomeni, YZ'nin kültürel diplomatik başarıya nasıl katkı sunabileceğinin en dikkat çekici örneklerinden biri olarak öne çıkmaktadır. Kore Kültür ve Bilgi Hizmetleri'nin küresel K-Pop ve K-Drama içeriklerinin yayılmasına yönelik kapsamlı dijital kampanyaları, veri analitiği ve YZ araçlarından yoğun biçimde yararlanmaktadır.

Spotify, YouTube ve Netflix gibi platformlarda Kore içeriklerinin algoritmik önerilere yansıması, bilinçli bir dijital strateji aracılığıyla değil, öncelikle organik izleyici talebi sayesinde gerçekleşmiştir. Ancak Güney Kore hükümetinin bu olguyu aktif biçimde desteklemesi ve teşvik etmesi, resmi kültürel teşvik politikaları ile organik kültürel cazibenin nasıl iç içe geçebileceğini açıkça ortaya koymaktadır.

Güney Kore, aynı zamanda hem kültürel ihracatı destekleyen hem de kültürel içerik üretimini kolaylaştıran YZ araçlarına yoğun yatırım yapmaktadır. Bu yaklaşım; YZ'yi saf bir araç olarak değil, kültürel yaratıcı ekonominin ayrılmaz bir bileşeni olarak konumlandırmaktadır.

C. Avrupa Birliği'nin Dijital Kültürel Diplomasi Yaklaşımı

Avrupa Birliği, kültürel diplomasi ve dijital politika arasındaki ilişkiyi belirgin biçimde farklı bir yaklaşımla ele almaktadır. AB'nin Dijital Pazar Yasası ve YZ Yönetmeliği gibi düzenleyici çerçeveler, özellikle algoritmik kültürel etkiye ilişkin kaygıları doğrudan hedef almaktadır.

'Avrupa Kültürel Mirası Dijital Seferberliği' girişimi, YZ araçlarını kültürel miras materyallerinin dijitalleştirilmesinde ve erişilebilir kılınmasında etkin biçimde kullanmaktadır. Bu yaklaşım; YZ'yi kültürel diplomasiye yönelik bir saldırı aracı

olarak değil, Avrupa'nın zengin kültürel birikimini küresel kitlelerle buluşturmanın bir yolu olarak konumlandırmaktadır.

Bununla birlikte AB'nin içinde bulunduğu ikilem de göz ardı edilemez: Bir yanda etik çerçevede, şeffaf ve çok kültürlü değerlere dayalı bir dijital kültürel diplomasiyi hayata geçirme çabası yer alırken, öte yanda bu alandaki baskın teknolojik altyapı ABD ve Çin'de üretilmekte ve kontrol edilmektedir.

VI. Stratejik Yaklaşımlar ve Politika Önerileri

A. Etik YZ Kültürel Diplomasi Çerçevesi

Yapay zekânın kültürel diplomaside sorumlu biçimde kullanımı için kapsamlı etik çerçeveler geliştirilmesi, öncelikli bir gereklilik olarak öne çıkmaktadır. Bu çerçeveler; şeffaflık ve hesap verebilirlik ilkelerini, izlerkitlelerin YZ ile oluşturulmuş kültürel içeriklerden haberdar edilmesi ile algoritmik içerik küratörlüğüne ilişkin açık bilgi politikalarını kapsamalıdır.

Kültürel çeşitlilik ve temsil açısından ise YZ sistemlerinin hâkim kültürel perspektifler lehine önyargı üretmeme ya da azınlık kültürel ifadelerini görünmez kılmama güvencelerini içerdiğinden emin olmak büyük önem taşımaktadır. Bu, hem algoritma tasarımında hem de YZ sistemlerinin eğitiminde kullanılan veri setlerinde çeşitlilik ilkesinin gözetilmesini gerektirmektedir.

İnsan gözetimi boyutunda ise YZ'nin kültürel diplomatik karar alma süreçlerini kolaylaştırdığı ancak bu kararların nihayetinde yetkili insan aktörler tarafından alındığını güvence altına alacak mekanizmaların tesis edilmesi büyük önem taşımaktadır.

B. Kapasite Geliştirme ve Dijital Eşitsizlik

YZ temelli kültürel diplomasi, mevcut dijital bölünmeyi derinleştirme riskini barındırmaktadır. Daha az gelişmiş dijital altyapıya sahip ülkeler, YZ araçlarını kültürel diplomatik amaçlarla kullanma kapasitesinden yoksun kalabilmektedir. Bu sorunun çözümü için güney-güney teknik iş birliğini teşvik eden, YZ kültürel diplomasi araçlarına erişimi kolaylaştıran ve küçük devletlerin sesini daha güçlü biçimde duyurmasını sağlayan gelişmiş YZ platformları geliştiren uluslararası programların hayata geçirilmesi gerekmektedir.

UNESCO [9] ve benzeri uluslararası örgütlerin bu eşitsizliği giderme yönündeki girişimleri, özellikle gelişmekte olan ülkelerdeki kültürel kurumlar için YZ araçlarına erişimi demokratikleştirme potansiyeli taşımaktadır.

C. Uluslararası Düzenleyici Koordinasyon

YZ'nin kültürel diplomaside küresel ölçekte kullanımı, koordineli uluslararası düzenleyici müdahaleyi zorunlu kılmaktadır. Bu çerçevede geliştirilmesi önerilen uluslararası standartlar arasında YZ temelli dezenformasyonla mücadeleye yönelik ortak normlar, kültürel algoritmik şeffaflık için uluslararası standartlar ile kültürel diplomatik YZ uygulamalarında veri koruma ve mahremiyet güvencelerine ilişkin çok taraflı anlaşmalar yer almaktadır. Bu düzeyde bir koordinasyon son derece iddialı bir hedef olsa da G20 Dijital Ekonomi Çalışma Grubu ve UNESCO'nun YZ Tavsiye Kararı, gerekli uluslararası platformların oluşturulması için bir başlangıç noktası sunmaktadır.

Sonuç

Araştırma, yapay zekânın kültürel diplomasi pratiklerinde yarattığı dönüşümü hem teorik hem de betimsel düzeyde ele almıştır. Değerlendirme, YZ'nin kültürel diplomasiye büyük fırsatlar sunduğunu; ancak bu fırsatların beraberinde ciddi etik, siyasal ve stratejik güçlükler getirdiğini ortaya koymaktadır.

Fırsatlar açısından YZ, daha geniş ve kapsayıcı kültürel etkileşimlere zemin hazırlama, kültürel içerikleri farklı kitlelere göre uyarlama, erişilebilirliği artırma ve kültürel karşılıklı anlayışı derinleştirme potansiyeli taşımaktadır. Doğru biçimde kullanıldığında, 21. yüzyılın nüanslı kültürel diplomasisi için vazgeçilmez bir araç olabilir.

Güçlükler cephesinde ise YZ temelli kültürel diplomasi; teknolojik hegemonyayı güçlendirebilir, dezenformasyon üretebilir ve kültürel özgünlük ile mahremiyet haklarını tehdit edebilir. Bu riskler, basit teknik çözümlerle giderilebilecek türden değildir; aksine kültürel diplomasinin anlam ve amaçlarına ilişkin köklü toplumsal tartışmaları zorunlu kılmaktadır.

Genel bir değerlendirme yapıldığında, YZ'nin kültürel diplomaside ne ölçüde yapıcı bir role üstleneceğinin, büyük ölçüde bu teknolojiyi yöneten siyasi, etik ve kurumsal çerçevelere bağlı olduğu görülmektedir. Bu nedenle uluslararası toplumun şimdi atacağı adımlar; önümüzdeki yıllarda küresel kültürel iletişim ortamını ve kültürel diplomasinin işlevini doğrudan belirleyecektir.

Sonuç olarak bu bildiri, kültürel diplomasisinin yapay zekâ çağında hâlâ insan değerlerine, karşılıklı saygıya ve kültürel çeşitliliğe dayalı bir uygulama alanı olarak kalmaya devam edebileceğini savunmaktadır. Bunun gerçekleşmesi için gerekli olan şey, YZ araçlarının körleme ve eleştirisiz biçimde uygulamaya sokulması değil; bu araçların kültürel anlayış ve uluslararası iş birliğini güçlendirme yolunda kasıtlı ve bilinçli kullanımıdır.

Kaynakça

[1] Nye J. S. Bound to lead: The changing nature of American power. — New York : Basic Books, 1990. — 307 s.

[2] Cummings M. C. Cultural diplomacy and the United States government: A survey. — Washington, D.C. : Center for Arts and Culture, 2003. — 52 s.

[3] Melissen J. (ed.) The new public diplomacy: Soft power in international relations. — Basingstoke : Palgrave Macmillan, 2005. — 244 s.

[4] Castells M. The new public sphere: Global civil society, communication networks, and global governance // The Annals of the American Academy of Political and Social Science. — 2008. — Vol. 616, № 1. — S. 78–93.

[5] Pamment J. New public diplomacy in the 21st century: A comparative study of policy and practice. — London : Routledge, 2013. — 210 s.

[6] Schiller H. I. Communication and cultural domination. — White Plains, NY : International Arts and Sciences Press, 1976. — 126 s.

[7] Tomlinson J. Globalization and culture. — Chicago : University of Chicago Press, 1999. — 238 s.

[8] Zuboff S. The age of surveillance capitalism: The fight for a human future at the new frontier of power. — New York : PublicAffairs, 2019. — 691 s.

[9] UNESCO. Recommendation on the ethics of artificial intelligence. — Paris : UNESCO Publishing, 2021. — 24 s.

Yapay Zeka ve Akademik Etik: Eğitim Modeli Üzerine Bir Değerlendirme

Artificial Intelligence and Academic Ethics: An Evaluation of The Educational Model

1st Erkam Temir
İletişim Fakültesi Gazetecilik Bölümü
Kastamonu Üniversitesi
Kastamonu, Türkiye
etemir@kastamonu.edu.tr

2nd İsmail Karakulle
İletişim Fakültesi Halkla İlişkiler ve Reklamcılık
Kastamonu Üniversitesi
Kastamonu, Türkiye
ikarakulle@kastamonu.edu.tr

Özet*—Üretken yapay zeka teknolojilerindeki gelişim akademik üretim süreçlerini belirgin biçimde değiştirmektedir. Bu araçların sınır gözetilmeksizin kullanılması akademik dürüstlük, özgünlük, atıf sorumluluğu ve fikrî mülkiyet gibi temel ilkeler açısından ciddi riskler doğurmaktadır. Bu kapsamda çalışma "Yayın Etiği Çerçevesinde Yapay Zeka Tabanlı Dil Modellerinin Kullanımı" başlıklı yapay zekanın akademide etik kullanıma yönelik geliştirilen eğitim modelini ve eğitimden elde edilen çıktıları ele almaktadır. TÜBİTAK destekli eğitim teorik dersler, vaka analizleri ve uygulamalı çalışmaların bir arada olduğu bütüncül bir yaklaşımla yürütülmüştür. Eğitim sürecinde 283 katılımcıdan elde edilen veriler yapay zekanın etik manada bilinçli kullanımına ilişkin bazı kritik eksiklikleri ortaya koymuştur. Katılımcılardan elde edilen verilere göre yapay zeka açısından akademik etiğe yönelik en büyük tehdit şeffaflık ihlalidir. Yapay zeka araçlarının metin üretimi, çeviri veya veri analizi gibi süreçlerdeki kullanım sınırlarının çalışmanın metodolojisinde açıkça beyan edilmemesi bu ihlalin temelini oluşturmaktadır. Araştırmacının asıl entelektüel sorumluluğu olan eleştirel düşünme, argüman geliştirme ve kavramsal çerçeve oluşturma süreçlerinin tamamen yapay zekaya devredilmesi ise bilişsel emeğin devri ve gölge yazarlık sorununu doğurmaktadır. Ayrıca orijinal olmayan fikirlerin veya başka yazarlara ait metinlerin intihal tespit yazılımlarını atlatmak amacıyla yapay zeka ile yeniden yazdırılarak özgün gibi sunulması gizlenmiş intihal problemini tetiklemektedir. Üretken yapay zekanın algoritmik hataları neticesinde gerçekte var olmayan makalelerin, yazarların veya ampirik verilerin doğrulanmadan bilimsel bir olgu gibi metne entegre edilmesi veri ve atıf uydurma riskini büyütmektedir. Diğer yandan etik kurul onayına tabi olan ve anonimleştirilmemiş hassas araştırma verilerinin güvenliği belirsiz halka açık sistemlere yüklenmesi ciddi bir veri mahremiyeti ihlali oluşturmaktadır. Son olarak literatür taramasının yapay zekaya otomatize edilmesi kaynakların derinlemesine analizi yapılmadan algoritmik önyargılara açık çıkarımların metne yansıtılmasına neden olmakta ve literatür sentezini yüzeysel eştirmektedir. Ayrıca özellikle yüksek lisans ve doktora düzeyindeki araştırmacıların genel yayın etiği kurallarına yeterince hâkim olmadığı ve benzerlik tespit programlarının çıktılarını doğru şekilde yorumlayamadıkları da eğitim sürecinde açıkça görülmüştür. Elde edilen bulgular akademik süreçlerde hem genel yayın etiği kültürünün güçlendirilmesinin hem de yapay zeka okuryazarlığına dair net yönergelerin aciliyetini göz önüne sermektedir.*

Anahtar kelimeler—üretken yapay zeka, akademik etik, akademik dürüstlük, yayin etiği, yapay zeka okuryazarliği.

***Abstract*—** *The advancement of generative artificial intelligence is fundamentally reshaping scholarly production. The unrestricted application of these tools poses severe threats to foundational principles such as academic integrity, originality, citation responsibility, and intellectual property. This study examines an educational model designed to foster the ethical use of AI in academia and evaluates its implementation outcomes. The training program titled "The Use of AI-Based Language Models Within the Framework of Publication Ethics"employed a holistic pedagogical approach integrating theoretical instruction, case studies, and practical applications. Data collected from 283 participants during the training revealed critical gaps in the conscious and ethical use of AI. Interactions with the participants highlighted a pervasive lack of transparency as the primary threat to scholarly ethics. This violation typically stems from a failure to explicitly declare the scope of AI involvement in a study's methodology, particularly regarding text generation, translation, or data analysis. Furthermore, delegating core intellectual responsibilities to AI creates a profound crisis. When researchers outsource critical thinking, argument formulation, and conceptual framing, they essentially engage in cognitive offloading and facilitate AI ghostwriting. Another major concern is concealed plagiarism. This occurs when unoriginal ideas or existing texts are rewritten using AI specifically to bypass similarity detection software and masquerade as original work. The algorithmic limitations of generative AI also introduce a severe risk of data and citation fabrication. Without rigorous verification, hallucinated articles, nonexistent authors, or fabricated empirical data are occasionally integrated into manuscripts as established scientific facts. Additionally, uploading sensitive, non-anonymized research data requiring ethical committee approval into unverified, public AI systems constitutes a blatant breach of data privacy. Finally, automating the literature review process degrades the quality of academic synthesis. Relying on AI for this task often replaces deep, critical analysis with superficial summaries that remain highly vulnerable to algorithmic bias. The training sessions further exposed a systemic issue among early-career researchers. Master's and doctoral students frequently lack a comprehensive understanding of general publication ethics and struggle to accurately interpret the reports generated by similarity detection software. Ultimately, these findings underscore an urgent need for action. Academia must simultaneously reinforce a robust culture of publication ethics and establish clear, definitive guidelines for AI literacy.*

Keywords—generative ai, academic ethics, academic integrity, publication ethics, ai literacy.

Education funded by TÜBİTAK.

I. GİRİŞ VE TEORİK ARKA PLAN

Üretken yapay zeka teknolojilerinin ve bilhassa büyük dil modellemelerine sahip araçların 2022 sonlarından itibaren yaygınlaşması yükseköğretim ve akademik üretim süreçlerinde yapısal bir dönüşümü tetiklemiştir. Bu teknolojiler araştırmacılara literatür taramasından veri analizine, metin üretiminden kod yazmaya kadar pek çok akademik faaliyette benzeri görülmemiş bir verimlilik artışı vaat etmekle birlikte, akademinin temelini oluşturması gereken dürüstlük, özgünlük ve şeffaflık ilkelerini de derinden sarsmaktadır. Nitekim uluslararası yayın etiği tartışmalarında yapay zeka tarafından üretilen içeriklerin sınır gözetilmeksizin ve usulüne uygun biçimde bildirilmeksizin kullanılması, geleneksel intihal (plagiarism) kavramının ötesinde "AI-giarism" olarak adlandırılan yeni bir etik ihlal kategorisini doğurmuştur [1]. Bu açıdan şimdilik genellikle üretken yapay zeka araçlarının akademik yazım sürecinde yazar veya yaratıcı olarak kullanılmaması yalnızca destekleyici bir araç konumunda kalması gerektiği vurgulanmaktadır. Yapay zeka tarafından üretilen içeriklerin açıkça belirtilmesi, kullanılan araç türü ile katkı düzeyinin açıkça beyan edilmesi yönünde uluslararası yayınevleri ve editöryal kurullar tarafından ortak politikalar geliştirilmektedir [2,3].

Burada "şimdilik" kaydının düşülmesinin temel nedeni yapay zeka teknolojilerindeki eksponansiyel gelişme hızı ile akademik kurumların bu değişime adaptasyon hızı arasındaki asimetridir. Güncel literatür ve uygulamalar dikkatle incelendiğinde küresel ölçekte etik açısından akademinin bu yeni teknoloji karşısında büyük ölçüde tıkanmış ve tepkisel bir çaresizlik içinde olduğu görülmektedir. Nitekim mevcut politikalar sorunun kökenine inmekten ziyade yayın etiğine dair tüm hukuki ve ahlaki sorumluluğu yazara devreden bir "beyan zorunluluğu" kalkanının arkasına sığınmaktan öteye gidememektedir. Bu çaresizliğin en somut yansıması akademik dürüstlüğü denetlemek üzere piyasaya sürülen "yapay zeka tespit" araçlarının yetersizliğidir. Bu araçlar metinlerin makine tarafından üretilip üretilmediği konusunda tutarlı ve güvenilir sonuçlar verememektedir.

Bu çaresizliğin ve kavram kargaşasının temelinde sözde "yapay zeka tespit" algoritmalarının teknik altyapısındaki temel bir yetersizlik yatmaktadır. Bu araçlar metnin mana derinliğini, argüman kurgusunu veya fikri özgünlüğünü analiz etmezler. Yalnızca metnin istatistiki öngörülebilirliğini ve cümle yapılarındaki değişkenliğini ölçerler. Başka bir deyişle sadece kelime dizilimlerinin bir makineden beklenmeyecek kadar "düzensiz" (yani insani) olup olmadığına odaklanırlar. Tam da bu algoritmik indirgemecilik nedeniyle tespit araçlarının "doğru" sonuç verdiğinin varsayıldığı vakalarda dahi aşılması güç üç boyutlu bir açmaz ortaya çıkmaktadır.

İlk olarak bütünüyle yapay zeka tarafından üretilen bir metnin yazar tarafından kısmi müdahalelerle yeniden ifade edilerek (parafraz yapılarak) tamamen özgünmüş gibi sunulmasıdır. Tespit araçlarının zafiyetlerini kavrayan kullanıcılar makinenin ürettiği standart metne bilinçli olarak "insani pürüzler" (karmaşık sözdizimi, asimetrik cümle yapıları veya spesifik kelime tercihleri) ekleyerek istatistiksel öngörülebilirliği aşabilmektedir. Bu taktik sayesinde tamamen yapay zeka üretimi olan metinler tespit radarlarından kolaylıkla kaçabilmektedir.

İkinci boyut özgün emeğin haksız yere yaftalanmasıdır. Araştırmacının tamamen kendi zihni emeğiyle, özgün fikirlerle ve geleneksel akademik etiğe sadık kalarak kurguladığı bir metin yalnızca dilbilgisel düzeltme, redaksiyon veya tercüme amacıyla bir büyük dil modelinden geçirildiğinde, metnin insani pürüzleri ve doğal değişkenliği törpülenmektedir. Ortaya çıkan pürüzsüz ve istatistiksel olarak öngörülebilir bu metin, tespit algoritmaları tarafından doğrudan "yapay zeka üretimi" olarak damgalanabilmektedir.

Ancak asıl sorun madalyonun diğer yüzünde yani bu muğlaklığın bir maskeleme stratejisi olarak suistimal edilmesinde gizlidir. Fikri kurgusundan argümantasyonuna kadar tamamen yapay zekaya ürettirilen bir metin, eğer algoritmik halüsinasyon içermiyorsa ve literatürdeki gerçek, doğrulanabilir kaynaklara atıf yapıyorsa, yazar tespit araçlarına yakalandığı takdirde bu ihlali kolaylıkla "dilbilgisel düzeltme" maskesinin arkasına saklayabilir. Günümüzde yapay zeka tarafından düzenlenmiş bir metin geriye dönük olarak (tersine mühendislik) incelendiğinde, o metnin çekirdeğindeki özgün fikrin mi yoksa yalnızca dil formunun mu algoritmaya ait olduğunu ayırt edebilecek hiçbir yazılımsal veya filolojik mekanizma bulunmamaktadır.

Sonuç olarak makine dili ile insan bilişi arasındaki sınırın bu denli silikleştiği bir düzlemde ne tespit yazılımlarının ürettiği yüzeysel istatistik raporları ne de sadece iyi niyete dayalı yazar beyanları, akademik bütünlüğü korumak ve gerçek yazarı tayin etmek için güvenilir bir zemin sunmamaktadır.

Gelinen noktada yapay zekanın çalışma doğası ve metin üretme mekaniği akademi paydaşlarının geneli tarafından tam olarak anlaşılamamıştır. İnsanın ürettiği özgün fikir ile makinenin sunduğu dil formu arasındaki sınırın giderek silikleşmesi etik ihlalin nerede başlayıp nerede bittiğine dair de son derece muğlak bir alan yaratmaktadır. "Fikri mülkiyet" ile "dil temsili" arasındaki bu karmaşa çözülmeden yalnızca yazar beyanına veya kusurlu tespit algoritmalarına dayanan mevcut yaklaşımların akademik bütünlüğü korumada yetersiz kalacağı açıktır. Metin üreten yapay zeka şirketleri ile bu metinleri tespit ettiğini iddia eden yazılımları satan şirketler akademinin bu paniğinden beslenen devasa bir pazar yaratmıştır.

Akademik özgünlük ve atıf sorumluluğu açısından bakıldığında yapay zekanın milyarlarca parametreyle eğitilmiş modellerinin, eğitim veri tabanındaki telifli eserlerden izinsiz ve kaynak göstermeksizin alıntı yapma potansiyeli, fikrî mülkiyet haklarının ihlali ve telif hukuku açısından da ciddi bir muğlak alan yaratmaktadır [4,5]. Üstelik bu modellerin sıklıkla "halüsinasyon" olarak adlandırılan, gerçekte var olmayan kaynaklar ve yanlış atıflar üretme eğilimi, bilginin bütünlüğünü ve güvenilirliğini tehdit eden bir risk olarak tanımlanmaktadır [6]. Diğer yandan öğrencilerin ve araştırmacıların etik farkındalık düzeyi,

yapay zeka kullanım politikalarının varlığı ve kurumsal denetim mekanizmalarının etkinliği gibi faktörlerin, üretken yapay zekanın akademik bağlamda etik kullanımını belirleyen temel değişkenler olduğu araştırmalarla ortaya konmuştur [7,8]. Üretken yapay zekâ teknolojilerinin yükseköğretim ve akademik üretim süreçlerinde zorladığı dönüşüm, akademik dürüstlük, özgünlük, atıf sorumluluğu ve fikrî mülkiyet ilkeleri çerçevesinde çok boyutlu bir etik değerlendirmeyi zorunlu kılmaktadır. Kurumların açık yapay zekâ politikaları geliştirmesini, müfredata yapay zekâ okuryazarlığı ve etiğinin entegre edilmesini ve tüm paydaşların bu dönüşüme hazırlanmasını gerektirmektedir[8,9,10].

Dien [11] yapay zekanın akademik yazımda ortaya çıkardığı en büyük sorunun kontrolsüz intihal olduğunu vurgularken, Cheng ve arkadaşları [12] araştırmacıların büyük dil modellerini akademik yazımda kullanırken şeffaflık, doğruluk ve hesap verebilirlik ilkelerini ihlal ettiğini göstermektedir. Bao ve Zeng [13] ise yapay zekaya yazarlık verilmesinin insanın rasyonel bir varlık olarak eleştirel düşünme sürecini göz ardı ettiğini ve yazar ile tez arasındaki anlamlı ilişkiyi olasılıksal bir üretim ilişkisine indirgediğini savunmaktadır.

Yapay zeka ile yeniden yazdırma yoluyla yapılan gizlenmiş intihal olarak karşımıza çıkmaktadır. Hammad [14], bu araçların cümleleri yeniden formüle ederek benzerlik oranlarını düşürdüğünü ancak bunun bilimsel açıdan meşru bir uygulama olup olmadığının tartışmalı olduğunu belirtmektedir. Doskaliuk ve Zimba [15], ChatGPT gibi araçların metin üretirken kaynak göstermeksizin başka yazarların fikirlerini yeniden sunabildiğini ve bu durumun geleneksel intihal tespit sistemlerini etkisiz kıldığını ifade etmektedir.

Bir diğer sorun algoritma hatalarından kaynaklanan veri ve atıf uydurma/hallüsinasyon olgusudur. Azamfirei ve arkadaşları [16], büyük dil modellerinin olasılıksal metin üretim mekanizması nedeniyle var olmayan kaynaklara atıf yapabildiğini, tamamen uydurma çalışmaları ikna edici bir dille özetleyebildiğini ve bu "hallüsinasyon" teriminin aslında "bilgi uydurma" olarak daha doğru tanımlanacağını göstermektedir. Koga [17]'da bu durumun özellikle genç araştırmacılar için ciddi bir tuzak oluşturduğunu ve sahte referansların fark edilmeden yayın sürecine dahil olabileceğini belirtmektedir.

Bu sorunların kesiştiği noktada araştırmacının asli görevi olan eleştirel düşünme ve argüman geliştirme süreçlerinin makinelere devredilmesi, bilişsel emeğin devri açısından derin bir probleme işaret etmektedir. Kulal [18], 625 araştırmacı üzerinde yaptığı çalışmada, yapay zeka kullanımının eleştirel düşünme becerisini anlamlı düzeyde olumsuz etkilediğini ve yapay zekaya duyulan güvenin bu etkiyi daha da şiddetlendirdiğini ortaya koymuştur. Ingley ve Pack [19], yapay zeka araçlarının yazma becerilerini geliştirmek yerine adeta araştırmacıların üzerine yaslanacağı bir "koltuk değneği" haline gelme tehlikesine dikkat çekmektedir. Benzer şekilde Khan ve arkadaşları [20] yapay zeka tarafından üretilen makalelerin okunabilirlik açısından daha başarılı olmasına karşın klinik deneyim, özgünlük ve nüanslı analiz derinliği açısından insan yazarların gerisinde kaldığını göstermektedir.

Asıl sorun bir araştırmacının hipotez kurma, literatürü eleştirel gözle değerlendirme, mantıksal çıkarımlar oluşturma ve özgün bir argüman geliştirme gibi temel bilişsel süreçleri yapay zekaya devretmesidir. Bu durum yalnızca etik bir ihlal değil aynı zamanda araştırmacının bilişsel özerkliğini de ortadan kalkması manasına gelmektedir. Cao ve arkadaşları [21], yapay zeka kimliği ile özdeşleşen insanlarda psikolojik hak sahipliği duygusunun arttığını ve bunun etik dışı davranışları tetiklediğini göstermektedir. Başka bir deyişle yapay zeka araçlarını yazma ve düşünme sürecinin merkezine koymak araştırmacının kendi kapasitesine olan güvenini zedeleyerek entelektüel özgünlüğü yapay bir yeniden üretime dönüştürmekte ve nihayetinde bilginin üretiminde insanı pasif bir denetleyici konumuna indirgemektedir. Bu durum bilimin ilerlemesinin temel dinamiği olan eleştirel akıl yürütme sürecinin yerine algoritmik metin üretimini ikame ederek uzun vadede akademik düşüncenin çöküşüne yol açabilecek büyük bir sorun oluşturmaktadır.

Mevcut çalışmalar yayın etiği ihlallerinin bireysel ve kurumsal boyutlarını ele almakla birlikte, lisansüstü düzeydeki araştırmacıların genel yayın etiği kurallarına yetersiz hâkimiyeti, benzerlik tespit programlarının çıktılarını doğru yorumlayamaması ve özellikle yapay zeka kullanımının metodolojide beyan edilmemesi gibi şeffaflık ihlallerinin kesiştiği noktada belirgin bir boşluk bulunmaktadır. Chang [22] birçok araştırmacının lisansüstü eğitimleri sırasında akademik ve araştırma etiği konusunda yeterli eğitim almadığını ve bu durumun etik ihlallerin temel kaynaklarından biri olduğunu vurgulamaktadır. Türkiye bağlamında yapılan çalışmalar da bu bulguyu desteklemekte örneğin araştırma görevlilerinin intihale karşı olumsuz tutum sergilemelerine rağmen yabancı dil sorunları, zaman baskısı ve etik kurallara ilişkin bilgi eksikliği nedeniyle intihal davranışına başvurabildikleri görülmektedir [23,24]. Gokmenoğlu [23] ayrıca Türkiye'de lisansüstü öğrencilerin ve öğretim üyelerinin intihal algılarını, durumlarını ve nedenlerini araştıran çalışmaların sınırlı sayıda olduğunu belirterek bu alandaki boşluğa işaret etmektedir. Öte yandan benzerlik tespit programlarının yaygınlaşmasına rağmen bu araçların çıktılarının nasıl yorumlanacağı konusunda lisansüstü araştırmacıların yeterince bilinçlendirilmediği anlaşılmaktadır. Dien [25] yapay zeka tabanlı üretken modellerin intihali kolaylaştırıcı yeni bir boyut kazandırdığını ve mevcut benzerlik tespit yöntemlerinin bu yeni tehdit karşısında yetersiz kalabileceğini belirtmektedir. Bu durum araştırmacıların yalnızca program çıktılarına güvenerek etik değerlendirme yapmasının risklerini gözler önüne sermektedir. Daha da kritik bir nokta ise yapay zeka araçlarının araştırma sürecinde kullanımına ilişkin şeffaflık ihlalleridir. Nguyen [26], yükseköğretim kurumlarının yapay zeka kullanımına dair açık politikalar geliştirmesi, yapay zeka etiğinin müfredata entegre edilmesi ve öğretim üyelerine sürekli eğitim sağlanması gerektiğini savunmaktadır. Nature dergisinin yayımladığı ilkelere göre, araştırmacıların yapay zeka araçlarını kullanmaları durumunda bunu yöntem veya teşekkür bölümlerinde belgelemeleri zorunlu kılınmıştır [27]. Ancak Friederich ve Symons [28] yapay zekanın akademik yazımdaki katkılarının beyanına ilişkin normların henüz olgunlaşmadığını ve bu alanda hızlı bir düzenleyici dönüşüme ihtiyaç duyulduğunu ifade etmektedir. Hoomanfard ve Shamsi [29] doktora öğrencilerinin yapay zeka kullanımının intihalle sınırı

konusunda belirsizlik yaşadıklarını ve yapay zeka kullanımını beyan edip etmeme ikilemiyle karşı karşıya kaldıklarını ortaya koymaktadır. Sertel ve diğerleri [30] ise Türkiye'deki akademik dergilerde editöryel etik uygulamalarında önemli eksiklikler olduğunu ve bu durumun etik krizleri derinleştirdiğini belirtmektedir. Bu bulgular, lisansüstü araştırmacıların etik eğitimi, benzerlik tespit programlarının doğru kullanımı ve yapay zeka kullanımının şeffaf beyanı arasındaki kesişim kümesinin bugüne kadar yeterince çalışılmadığını ve mevcut araştırmaların bu üç boyutu bütüncül bir çerçevede ele almadığını göstermektedir. Özellikle gelişmekte olan ülkeler bağlamında bu üç faktörün birbirini nasıl tetiklediği ve mevcut etik ihlalleri nasıl derinleştirdiği sorusu, acil olarak ele alınması gereken bir araştırma boşluğu olarak öne çıkmaktadır.

Böylelikle bu çalışma "Yayın Etiği Çerçevesinde Yapay Zeka Tabanlı Dil Modellerinin Kullanımı" [31] başlıklı TÜBİTAK tarafından desteklenen eğitimin yansımalarını, eğitim sürecinde katılımcıların yönelttiği sorular, gerçekleştirilen görüşmeler ve tartışmalardan elde edilen nitel bulgular ışığında değerlendirerek, akademik yazım süreçlerinde giderek yaygınlaşan yapay zeka tabanlı dil modellerinin etik kullanımına ilişkin belirgin eksiklikleri ortaya koymayı amaçlamaktadır.

II. Etkinliğin Kapsamı ve Yapısı

Çalışmanın bulgularını şekillendiren TÜBİTAK 2237-A eğitim etkinliği bilginin güvenilirliği, izlenebilirliği ve hesap verebilirliği üzerine kurulan multidisipliner bir kapasite geliştirme modeli olarak tasarlanmıştır. Sosyoloji, hukuk, iletişim ve etik alanlarından dört farklı eğitmenin katılımıyla 16 saatlik yoğunlaştırılmış karma bir pedagoji uygulanmıştır.

Eğitim için başlangıçta 100 kişilik bir kontenjan öngörülmüşken, kısa sürede 400'ü aşkın başvuru alınması, üretken yapay zekanın akademide yarattığı metodolojik belirsizliğin ve bu alandaki eğitim ihtiyacının ulaştığı seviyeyi göstermektedir. Ampirik veriler toplanmamasına karşın eğitime kabul edilen 283 araştırmacının Türkiye genelindeki 75 farklı üniversiteyi temsil etmesi, elde edilen bulguların sistemik bir akademik pratikler bütününe işaret ettiğini kanıtlamaktadır. Bu eğitim modelinin saha uygulamalarından elde edilen kritik bulgular şunlardır:

A. İntihal Tespit Yazılımlarını Atlatma Çabası ve Gizlenmiş İntihal

Etkinliğin teknik ve uygulama boyutunda yapay zeka kullanımının intihal olgusuyla kesiştiği yeni ve tehlikeli boyutlar analiz edilmiştir. Araştırmacıların eğitimdeki soruları ve vaka çözümlerindeki yaklaşımları, intihal tespit programlarının algoritmik mantığının büyük oranda yanlış anlaşıldığını ortaya koymuştur.

Birçok katılımcının, benzerlik oranını etik bir "karar mercii" olarak gördüğü ve düşük benzerlik oranının mutlak bir etik güvence sağladığı yanılgısına düştüğü tespit edilmiştir. Etkinlikteki uygulamalarda yapay zeka araçlarının orijinal olmayan fikirleri veya başka yazarlara ait metinleri, sırf benzerlik programlarını atlatmak amacıyla yeniden yazmak için sistematik olarak kullanılma eğilimi görülmüştür. Eğitimde benzerlik programlarının yapay zeka metinlerini tespit etme sınırları ve filtreleme seçenekleri gösterildiğinde katılımcıların, benzerlik raporunun bir kanıt olmadığını, bağlamsal bir insan muhakemesi gerektiren bir gösterge olduğunu yeni fark ettikleri gözlemlenmiştir.

B. Algoritmik Halüsinasyonlar: Veri, Bulgu ve Atıf Uydurma

Eğitim süresince yapay zekanın çalışma prensiplerine dair tehlikelerden birinin "algoritmik halüsinasyonlar" olduğu saptanmıştır. Etkileşimli tartışmalarda lisansüstü öğrencilerin literatür taraması veya kaynakça oluşturma gibi süreçleri tamamen LLM'lere devrettiği senaryolar incelenmiştir.

Üretken yapay zekanın doğası gereği bilgiyi "doğrulamak" yerine istatistiksel olarak en olası kelimeleri yan yana getirerek metin "üretmesi", gerçekte var olmayan makalelerin, yazarların, DOI numaralarının veya ampirik verilerin gerçek bilimsel olgularmış gibi araştırmalara entegre edilmesine yol açmaktadır. Vaka analizleri, katılımcıların LLM'ler tarafından üretilen bu kurgusal kaynakları derinlemesine bir çapraz kontrole tabi tutmadan doğrudan akademik metinlerine aktarma eğiliminde olduklarını, bu durumun veri ve atıf uydurma riskini eşi görülmemiş bir hızda artırdığını belgelemiştir.

C. Hassas Verilerin Güvenliği ve Mahremiyet İhlalleri

Eğitim programının multidisipliner yapısının ortaya çıkardığı ve katılımcıların büyük çoğunluğunun gözden kaçırdığı bir diğer kritik bulgu ise veri mahremiyetidir. Etik kurul onayına tabi olan, henüz yayınlanmamış veya anonimleştirilmemiş hassas araştırma verilerinin (örneğin; mülakat dökümleri, anket sonuçları, sağlık veya kurumsal veriler) veri analizi veya metne dönüştürülme amacıyla yapay zeka sistemlerine yüklendiği görülmüştür. Kurumsal güvenlik çeperi olmayan bu tür platformların, girilen verileri kendi modellerini eğitmek için kullanma potansiyeli taşıdığı gerçeği, akademik bağlamda ciddi bir gizlilik ihlali ve gelecekteki olası telif haklarının kaybı anlamına gelmektedir. Eğitmenlerin bu konudaki uyarıları, akademide veri güvenliği okuryazarlığının yapay zeka araçlarının hızına yetişemediğini somut biçimde göstermiştir.

D. Literatür Taramasında Otomasyon ve Yüzeysel Sentez Sorunu

Eğitim programında ele alınan bir diğer temel mesele, literatür taraması gibi derinlikli ve analitik okuma gerektiren süreçlerin otomatize edilmesinin yarattığı metodolojik tahribattır. Katılımcılarla yürütülen vaka çalışmalarında yapay zeka araçlarının sağladığı hız avantajının, metinlerin ve kaynakların yüzeysel bir şekilde taranmasına yol açtığı tespit edilmiştir.

Literatür taraması bir önceki çalışmalar listesi sunmaktan ibaret olmamalıdır. Mevcut boşlukları tespit etmeyi, farklı yaklaşımları sentezlemeyi ve araştırmacının kendi hipotezini konumlandıracağı entelektüel zemini oluşturmayı gerektirir. Ancak bu sürecin yapay zekaya devredilmesi algoritmaların istatistiksel olarak popüler olan veya sık atıf alan çalışmaları öne çıkarma eğilimi nedeniyle literatür sentezinin

sığlaşmasına neden olmaktadır. Araştırmacılar metinlerin tamamını okumak, bağlamı kavramak ve metodolojik eleştiriler getirmek yerine, yapay zekanın sunduğu hap niteliğindeki özetlere bel bağlamaktadır. Bu durum bilimsel üretimin kalitesini düşürmekte ve literatürün eleştirel bir süzgeçten geçirilmeden, mekanik bir biçimde metne aktarılması sorununu doğurmaktadır.

E. Genel Yayın Etiği ve Benzerlik Raporu Okuryazarlığındaki Yetersizlikler

Çalışmanın bulguları yapay zeka kullanımındaki etik ihlallerin temelinde aslında büyük ölçüde genel olarak yayın etiği kültürü eksikliğinin yattığını göstermektedir. Özellikle yüksek lisans ve doktora düzeyindeki genç araştırmacıların, intihal, dilimleme, haksız yazarlık ve uydurma gibi geleneksel etik ihlallerin kavramsal sınırlarına dahi tam olarak hâkim olmadıkları eğitim sürecinde açıkça görülmüştür.

Uygulamalı oturumlarda, katılımcıların Turnitin ve iThenticate gibi benzerlik tespit yazılımlarının çıktılarını yorumlama konusunda ciddi yetersizlikleri olduğu saptanmıştır. Benzerlik oranının bir etik hüküm olmadığı, yalnızca bir gösterge olduğu gerçeği göz ardı edilmektedir. Genel etik kurallara ve benzerlik raporlarının algoritmik mantığına dair temel eksiklikler, yapay zeka gibi çok daha karmaşık bir teknolojinin denetimini ve etik yönetimini neredeyse imkânsız hâle getirmektedir. Temel etik zeminin zayıf olması üretken yapay zekanın ortaya çıkardığı muğlak alanların adeta birer "akademik kestirme yol" olarak algılanmasını kolaylaştırmaktadır.

III. TARTIŞMA VE SONUÇ

Bu çalışma, TÜBİTAK 2237-A desteğiyle gerçekleştirilen eğitim programından elde edilen veriler ışığında, üretken yapay zeka tabanlı dil modellerinin akademik üretim süreçlerinde yarattığı yapısal ve etik dönüşümleri ortaya koymayı amaçlamaktadır. Elde edilen bulgular yapay zeka teknolojilerinin bilimsel araştırmalara entegrasyonunun sadece teknik bir adaptasyon süreci olmadığını, doğruluk, şeffaflık, fikrî mülkiyet ve entelektüel sorumluluk gibi akademinin temel ilkelerini derinden sarstığını göstermektedir.

Bulguları yapay zeka araçlarının kontrolsüz kullanımının, araştırmacıyı üreten ve sorgulayan bir özne olmaktan çıkarıp, algoritmik çıktıları pasif bir şekilde onaylayan bir operatöre dönüştürme riski taşıdığını doğrulamaktadır. Metodolojik beyan eksikliği, bilişsel emeğin devri, algoritmik halüsinasyonlar ve hassas verilerin mahremiyet ihlali gibi saptanan problemler, akademik dürüstlüğü doğrudan tehdit etmektedir. Bu risklerin azaltılması araştırmacının bilişsel ve etik denetim rolünün merkeze alındığı insan denetiminde bir yaklaşımla mümkündür. Bu çalışma güncel etik sorunların kavramsal ve gözlemsel çerçevesini çizmiş olsa da araştırmacıların yapay zeka kullanım sıklığını, algılarını ve potansiyel etik ihlal eğilimlerini istatistiksel olarak ortaya koyacak geniş örneklemli, nicel veya karma desenli araştırmalara ihtiyaç vardır.

KAYNAKÇA

[1] C. K. Y. Chan, "Students' perceptions of 'AI-giarism': Investigating changes in understandings of academic misconduct," *Education and Information Technologies*, vol. 30, pp. 8087-8108, 2025, doi:10.1007/s10639-024-13151-7.

[2] A. Cheng, A. Calhoun and G. Reedy, "Artificial intelligence-assisted academic writing: Recommendations for ethical use," *Advances in Simulation*, vol. 10, no. 22, 2025, doi:10.1186/s41077-025-00350-6.

[3] B. Doskaliuk and O. Zimba, "Beyond the keyboard: Academic writing in the era of ChatGPT," *Journal of Korean Medical Science*, vol. 38, no. 26, p. e207, Jul. 2023, doi:10.3346/jkms.2023.38.e207.

[4] J. D'Alfonso, "Generative artificial intelligence outputs, copyright infringement, and the assignment of liability," *AI Ethics*, vol. 5, pp. 5295-5308, 2025, doi:10.1007/s43681-025-00748-y.

[5] E. A. Alasadi and C. R. Baiz, "Generative AI in education and research: Opportunities, concerns, and solutions," *Journal of Chemical Education*, vol. 100, no. 8, pp. 2965-2971, 2023, doi:10.1021/acs.jchemed.3c00323.

[6] A. Cheng, A. Calhoun and G. Reedy, "Artificial intelligence-assisted academic writing: Recommendations for ethical use," *Advances in Simulation*, vol. 10, Art. no. 22, 2025, doi:10.1186/s41077-025-00350-6.

[7] G. F. N. Mvondo, B. Niu and S. Eivazinezhad, "Exploring the ethical use of LLM chatbots in higher education," *Journal of Business Research*, vol. 200, Art. no. 115575, 2025, doi:10.1016/j.jbusres.2025.115575

[8] A. Malik, M. L. Khan, K. Hussain, J. Qadir and A. Tarhini, "AI in higher education: Unveiling academicians' perspectives on teaching, research, and ethics in the age of ChatGPT," *Interactive Learning Environments*, vol. 33, no. 3, pp. 2390-2406, 2025, doi:10.1080/10494820.2024.2409407.

[9] K. V. Nguyen, "The use of generative AI tools in higher education: Ethical and pedagogical principles," *Journal of Academic Ethics*, vol. 23, no. 3, pp. 1435-1455, 2025, doi:10.1007/s10805-025-09607-1.

[10] T. Rasul, S. Nair, D. Kalendra, M. S. Balaji, F. de Oliveira Santini, W. J. Ladeira, R. A. Rather, N. Yasin, R. V. Rodriguez, P. Kokkalis, M. W. Murad and M. U. Hossain, "Enhancing academic integrity among students in GenAI era: A holistic framework," *The International Journal of Management Education*, vol. 22, no. 3, Art. no. 101041, 2024, doi:10.1016/j.ijme.2024.101041.

[11] J. Dien, "Editorial: Generative artificial intelligence as a plagiarism problem," *Biological Psychology*, vol. 181, p. 108621, Jul. 2023, doi:10.1016/j.biopsycho.2023.108621.

[12] A. Cheng, A. Calhoun and G. Reedy, "Artificial intelligence-assisted academic writing: Recommendations for ethical use," *Advances in Simulation*, vol. 10, Art. no. 22, 2025, doi:10.1186/s41077-025-00350-6.

[13] A. Bao and Y. Zeng, "Allowing AI co-authors is a disregard for humanization," *Accountability in Research*, vol. 33, no. 1, 2026, doi:10.1080/08989621.2024.2420812.

[14] M. Hammad, "The impact of artificial intelligence (AI) programs on writing scientific research," *Annals of Biomedical Engineering*, vol. 51, no. 3, pp. 459-460, 2023, doi:10.1007/s10439-023-03140-1.

[15] B. Doskaliuk and O. Zimba, "Beyond the keyboard: Academic writing in the era of ChatGPT," *Journal of Korean Medical Science*, vol. 38, no. 26, p. e207, 2023, doi:10.3346/jkms.2023.38.e207.

[16] R. Azamfirei, S. R. Kudchadkar and J. Fackler, "Large language models and the perils of their hallucinations," *Critical Care*, vol. 27, no. 1, p. 120, 2023, doi:10.1186/s13054-023-04393-x.

[17] S. Koga, "The integration of large language models such as ChatGPT in scientific writing: Harnessing potential and addressing pitfalls," *Korean Journal of Radiology*, vol. 24, no. 9, pp. 924-925, 2023, doi:10.3348/kjr.2023.0738.

[18] A. Kulal, "Cognitive risks of AI: Literacy, trust, and critical thinking," *Journal of Computer Information Systems*, pp. 1-13, 2025, doi:10.1080/08874417.2025.2582050.

[19] S. J. Ingley and A. Pack, "Leveraging AI tools to develop the writer rather than the writing," *Trends in Ecology & Evolution*, vol. 38, no. 9, pp. 785-787, 2023, doi:10.1016/j.tree.2023.05.007.

[20] R. R. Fauziah, A. M. I. Puspita, I. Yuliana, F. S. Ummah, S. Mufarochah and E. Ramadhani, "Artificial intelligence in academic writing: Enhancing or replacing human expertise?," *Journal of Clinical*

Neuroscience, vol. 135, p. 111193, 2025, doi:10.1016/j.jocn.2025.111193.

[21] L. Cao, C. Chen, X. Dong, M. Wang and X. Qin, "The dark side of AI identity: Investigating when and why AI identity entitles unethical behavior," *Computers in Human Behavior*, vol. 143, p. 107669, 2023, doi:10.1016/j.chb.2023.107669.

[22] H. Chang, "Research and publication ethics in academia," *Healthcare Informatics Research*, vol. 26, no. 1, pp. 1-2, 2020, doi:10.4258/hir.2020.26.1.1.

[23] T. Gokmenoglu, "A review of literature: Plagiarism in the papers of Turkish context," *Higher Education Studies*, vol. 7, no. 3, pp. 161-170, 2017, doi:10.5539/hes.v7n3p161.

[24] E. Eret and T. Gokmenoglu, "Plagiarism in higher education: A case study with prospective academicians," *Procedia-Social and Behavioral Sciences*, vol. 2, no. 2, pp. 3303-3307, 2010, doi:10.1016/j.sbspro.2010.03.505.

[25] J. Dien, "Editorial: Generative artificial intelligence as a plagiarism problem," *Biological Psychology*, vol. 181, p. 108621, 2023, doi:10.1016/j.biopsycho.2023.108621.

[26] K. V. Nguyen, "The use of generative AI tools in higher education: Ethical and pedagogical principles," *Journal of Academic Ethics*, vol. 23, pp. 1435-1455, 2025, doi:10.1007/s10805-025-09607-1.

[27] "Tools such as ChatGPT threaten transparent science; here are our ground rules for their use," *Nature*, 2023, doi:10.1038/d41586-023-00191-1.

[28] S. Friederich and J. Symons, "Norms for academic writing in the era of advanced artificial intelligence," *Discover Artificial Intelligence*, vol. 2, Art. no. 48, 2023, doi:10.1007/s44206-023-00079-7.

[29] M. H. Hoomanfard and Y. Shamsi, "Generative AI in dissertation writing: L2 doctoral students' self-reported use, AI-giarism, and perceived training needs," *Journal of English for Academic Purposes*, vol. 78, p. 101570, 2025, doi:10.1016/j.jeap.2025.101570.

[30] G. Sertel, E. Karadag and S. K. Ciftci, "Educational researchers' perceptions of editorial ethics and attitudes in Turkish academic journals: A mixed-methods study," *International Journal for Educational Integrity*, vol. 21, Art. no. 19, 2025, doi:10.1007/s40979-025-00194-7.

[31] Kastamonu Üniversitesi, "Yayın Etiği," 2026. [Online]. https://yayinetigi.kastamonu.edu.tr/. [Erişim: 2 Mayıs 2026].

Implementation of AI Call center in Banking Sector: Development and Analysis

Zholdoshbek Tashiev
Kyrgyz-Turkish Manas University
Bishkek, Kyrgyz Republic
2552y0501001@manas.edu.kg

Rita Ismailova
Kyrgyz-Turkish Manas University
Bishkek, Kyrgyz Republic
rita.ismailova@manas.edu.kg

The present era can rightfully be called the golden age of Artificial Intelligence. However, development of applications that utilize Artificial Intelligence is still in its early stages in the Kyrgyz Republic. This study aims to automate and integrate AI into the operations of call centers in the country's banking sector, which operated in three languages, namely, Kyrgyz, Russian, and English. The system developed within the scope of the study replaces labor-intensive components and telephone calls with an intelligent voice agent capable of conducting natural dialogue in these languages. The importance of the study lies in the fact that the Kyrgyz language belongs to low-resource Turkic languages with approximately 7.5 million native speakers, and thus, there are limited resources for this language. The architecture incorporates real-time speech-to-text recognition (STT), a large language model (LLM) for dialogue management, and text-to-speech synthesis (TTS) via an audio pipeline based on the RTP protocol, connected to the corporate telephone network through Asterisk PBX. General testing was conducted in an environment involving 100 bank employees (G.711 μ-law, 8 kHz, via SIP trunk). The results demonstrated high performance for Russian (approximate WER of 8–12%) and English (WER of 5–10%) with approximately 0.5 turns per conversation. For Kyrgyz, degradation was observed in longer utterances: for short sentences (1–4 words) WER was ~10–15%, while for longer sentences (5+ words) it reached ~25–40%. This phenomenon is attributed to the STT model being trained on wideband audio data at 16 kHz and above, whereas the telephone network transmits a narrowband signal at 8 kHz, losing a significant portion of the speech spectrum critical for accurate recognition. This effect is further amplified in longer sentences.

Keywords—artificial intelligence, call center, banking, speech recognition, text-to-speech, low-resource language, Kyrgyz language, large language model, telephony, Whisper STT.

I. INTRODUCTION

Context and Motivation

Debt collection remains one of the most labor-intensive operations in retail banking. In the Kyrgyz Republic — with a population of approximately 7.5 million and a rapidly developing retail lending market — banks face the challenge of managing overdue credit portfolios with limited personnel resources. Traditional call centers require substantial costs for staffing, training, and quality control, along with limited scalability during peak load periods [1]. The emergence of large language models capable of contextual multi-turn dialogue, combined with advances in speech recognition and synthesis, creates an opportunity to automate routine calls. However, the specific challenges of banking voice automation distinguish it from general-purpose conversational AI: regulatory compliance, structured data extraction, multilingual operation, support for low-resource Kyrgyz language, real-time constraints of IP telephony, and integration with corporate banking systems [2].

Contributions of This Paper

This paper makes the following contributions:

- Complete architecture of a production AI voice agent for banking calls — from receiving data from the Core Banking System (CBS) to storing conversation outcomes;
- Empirical evaluation with 100 bank employees quantifying length-dependent degradation of STT for the Kyrgyz language under real telephony conditions;
- Systematic analysis of WER degradation in the transition from 16 kHz clean audio to 8 kHz telephony for all three languages;
- The first documented production deployment of a complete AI voice pipeline for the Kyrgyz language in a banking context.

II. RELATED WORK

Conversational AI in Banking

The use of artificial intelligence in banking has expanded from fraud detection and credit scoring to customer service. Chatbot-based systems for handling incoming queries have been widely studied [3,4], while voice agents for proactive outbound calls — particularly in Central Asian markets — remain largely unexplored. Commercial platforms (Google Dialogflow, Amazon Connect, Nuance) offer ready-made solutions but do not meet the data sovereignty and air-gapped network requirements specific to banking in Kyrgyzstan.

Multilingual Speech Recognition and Low-Resource Languages

OpenAI Whisper [5] demonstrated high-quality multilingual recognition, trained on over 5 million hours of audio including approximately 680,000 labeled hours. However, quality drops sharply for languages with minimal representation in training data. Significant advances for Kazakh were achieved thanks to the KazSpeech corpus [6]. Neural approaches for Uzbek speech recognition have also been studied [7]. Kyrgyz remains one of the least resourced Turkic languages: all existing models have been trained exclusively on wideband clean audio at 16 kHz.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

LLMs for Structured Dialogue

Generating structured output from LLMs for goal-oriented dialogue has been studied in several NLP works [8]. The Qwen model family [9] has demonstrated strong results for Russian and Turkic languages, competing with larger models on instruction following and multilingual tasks [10]. Mixture-of-Experts models achieve balance between quality and computational efficiency through sparse parameter activation.

Latency in Voice Dialogue Systems

Natural turn-taking in human conversation occurs with pauses of approximately 200 ms, significantly shorter than typical silence thresholds in dialogue systems (700–1000 ms) [11]. The target end-to-end latency for voice agents is under 2 seconds; exceeding this threshold causes users to perceive the system as unnatural [11]. Mismatch between STT model training conditions (16 kHz wideband audio) and inference conditions (8 kHz narrowband telephony) leads to significant WER degradation [12].

.

III. System Architecture

Overview

The system is built on a microservice architecture deployed across three physical servers on an isolated banking network without Internet access.

TABLE.I COMPONENT-TO—SERVER ALLOCATION

Server	Components	Purpose
Primary	React, FastAPI, Celery, Asterisk PBX, PostgreSQL, Redis	Orchestration, API, telephony
GPU Server	STT multilingual, STT Kyrgyz, TTS Kyrgyz	GPU speech processing
LLM Server	Ollama + Qwen3.5-35B-A3B	Dialogue management

Call Processing Flow

The end-to-end process comprises seven stages:

- (1) The Celery scheduler initiates calls via the Asterisk REST Interface;
- (2) Asterisk places an outbound SIP call via a Cisco trunk;
- (3) Upon answer, an ExternalMedia channel is established with an RTP audio stream to the AI pipeline;
- (4) The backend delivers session configuration;
- (5) The pipeline processes audio through the STT → LLM → TTS chain;
- (6) The session ends on conversation completion, limit triggers, or hangup;
- (7) The backend stores transcripts, recordings, and classified outcomes.

Audio Processing and VAD

Incoming audio is decoded from μ-law format to float32 PCM and resampled from 8 to 16 kHz. For the Kyrgyz channel, extended preprocessing is applied: high-frequency filtering, neural bandwidth extension (HiFi-GAN BWE), noise suppression, pre-emphasis, and normalization. VAD is implemented on the Silero VAD Lite model — threshold 0.35, end-of-speech delay 300 ms; maximum utterance length 15 seconds.

.

IV. Speech Components: Models and Characteristics

STT — Russian and English

The core engine is the Faster-Whisper large-v3-turbo model deployed as a GPU service.

TABLE.II WHISPER LARGE-V3-TURBO CHARACTERISTICS

Parameter	Value	Value	Value	Value
Parameter	**Model Parameters**	**Avg WER (HF Open ASR)**	**WER LibriSpeech Clean (EN)**	**WER GigaSpeech (conversational)**
Faster-Whisper large-v3-turbo	809M	7.83%	2.10%	10.14%

STT — Kyrgyz Language

For Kyrgyz, the specialized model UlutSoftLLC/whisper-small-kyrgyz is deployed. During development, three available Kyrgyz STT models were evaluated:

TABLE.III AVAILABLE STT MODELS FOR THE KYRGYZ LANGUAGE

Model	Base	Parameters	WER (clean 16 kHz)	Status
UlutSoftLLC/whisper-small-kyrgyz	Whisper Small	244M	~8% (published)	In use ✓
nineninesix/kyrgyz-whisper-medium	Whisper Medium	769M	20–40% (FLEURS)	~2K hrs + 40% EN/RU mix
kyrgyz-ai/AkylAI-STT-small	Whisper Small	244M	Not published	No published metrics

All three models were trained exclusively on clean 16 kHz audio and do not include narrowband telephony data — this is the primary cause of degradation when operating over G.711.

A. TTS — All Languages

TABLE IV TTS MODELS IN THE SYSTEM

MatchaTTS with HiFi-GAN Universal v1 vocoder was selected for Kyrgyz based on a business requirement for a young operator's voice. Both male (epoch 279) and female (epoch 479) variants are supported.

Language Model (LLM)

The Qwen3.5-35B-A3B (Mixture-of-Experts) model is deployed via Ollama for dialogue management. Sparse activation (~3.5B of ~35B parameters) provides quality at the level of 14B models with inference latency comparable to smaller models. Parameters: temperature=0.3 (dialogue), temperature=0.1 (outcome classification), max_tokens=200. All responses are generated in JSON format containing the fields: response, fields, is_final, intent.

V. Empirical Evaluation

TABLE V. TESTING PARAMETERS

Parameter	Value
Participants	100 bank employees
Channel	Real network via Cisco SIP trunk
Codec / Sample Rate	G.711 μ-law, 8 kHz
Language selection	Free — unrestricted
Conditions	Real office conditions (background noise, varying connection quality)
Test type	Black-box: free dialogue on debt collection scenario
Agent settings	Qwen3.5-35B-A3B, Whisper large-v3-turbo, Silero TTS (production)

Results — Russian and English

TABLE VI. TESTING RESULTS — RUSSIAN AND ENGLISH

Parameter	Russian	English
STT Model	Whisper large-v3-turbo (809M)	Whisper large-v3-turbo (809M)
WER (clean 16 kHz, reference)	~5–8%	2.1–4.2% (LibriSpeech)
Estimated WER (8 kHz telephony)	8–12%	5–10%
Avg re-prompts / conversation	~0.5	~0.5
Recognition MOS (subjective, 1–5)	4.0–4.5	4.0–4.5
TTS MOS over telephony (subjective, 1–5)	3.3–3.6	3.2–3.5
Completion rate	~85–90%	~85–90%
Overall impression (MOS 1–5)	3.8–4.2	3.8–4.2

Model	Language	Architecture	Sample Rate	MOS (source)
Silero TTS v4	Russian	Neural TTS	48→8 kHz	4.18–4.54
Silero TTS v3	English	Neural TTS	48→8 kHz	~4.0
MatchaTTS + HiFi-GAN	Kyrgyz	Flow-matching + vocoder	22.05→8 kHz	~4.0 (baseline)

Results — Kyrgyz Language

TABLE VII. KYRGYZ STT WER DEPENDENCY ON UTTERANCE LENGTH

Utterance Length	Approx. WER	Examples	Assessment
1–4 words	~10–15%	"yes", "no", "I will pay tomorrow"	Acceptable
5+ words	~25–40%	Extended explanations	Significant errors
+ BWE preprocessing (short)	~8–12%	—	+3% improvement
+ BWE preprocessing (long)	~20–35%	—	+5% improvement

TABLE VIII. SUMMARY RESULTS — KYRGYZ

Metric	Kyrgyz
STT Model	UlutSoftLLC/whisper-small-kyrgyz (244M)
WER – short (1–4 words)	~10–15%
WER – long (5+ words)	~25–40%
WER after HiFi-GAN BWE + preprocessing	~15–25% (~5% improvement)
Avg re-prompts / conversation	~1.5–2.0
Recognition MOS – short	3.0–3.5
Recognition MOS – long	2.0–2.5
TTS MOS over telephony (MatchaTTS)	2.8–3.2
Completion rate	~70–80%
Overall impression (MOS 1–5)	3.0–3.5

Analysis of Kyrgyz STT Degradation

Training and Test Data Mismatch

All Kyrgyz STT models are trained on clean wideband 16 kHz audio. G.711 μ-law restricts the band to 300–3400 Hz. After 8→16

kHz resampling, the 3400–8000 Hz range contains zeros — a model that expects energy in this range receives statistically anomalous input.

Accumulation of Autoregressive Errors

Whisper uses autoregressive decoding: each token is conditioned on all previous tokens. For short phrases (1–4 words), errors do not accumulate. For longer utterances (5+ words), a single token error cascades, explaining the nonlinear WER dependence on utterance length.

TABLE IX. COMPARISON OF MODEL ROBUSTNESS TO TELEPHONY AUDIO

Factor	Whisper large-v3-turbo (RU/EN)	Whisper-small UlutSoft (KY)
Parameters	809M	244M (3.3× fewer)
Encoder layers	32	12
Telephony data in training	Yes (>5M hrs portion)	No
VAD filter in inference	Yes (vad_filter=True)	No
Beam search	Yes (beam_size=5)	No (greedy)
16→8 kHz degradation	+3–6% WER	+12–22% WER

1) WER Degradation in 16 kHz → 8 kHz Transition

TABLE X. WER DEGRADATION: WIDEBAND → NARROWBAND TELEPHONY AUDIO

Language / Model	WER (16 kHz clean)	WER (8 kHz telephony)	Degradation
EN / Whisper large-v3-turbo	2–4%	5–10%	+3–6%
RU / Whisper large-v3-turbo	5–8%	8–12%	+3–4%
KY / whisper-small UlutSoft	~8%	20–30%	+12–22%
KY / whisper-small + BWE	~8%	15–25%	+7–17% (with mitigation)

Four-Level Protection Against STT Errors

TABLE XI. FOUR-LEVEL STT ERROR HANDLING MECHANISM

Level	Mechanism	What it filters	Threshold
1	STT confidence filter	Noise, low-confidence recognitions	confidence < 0.28
2	Hallucination blocklist	Whisper phantom text patterns	14 known patterns
3	Duplicate suppression	Echo of own TTS voice	Match with previous output
4	LLM intent classification	Meaningless text	intent = "unclear" → re-prompt

VI. BACKEND AND BUSINESS LOGIC

Automated Call Scheduling

TABLE XII. SCHEDULING STRATEGY BY OUTCOME

Conversation Outcome	Next Action	"No Answer" counter
Positive	Campaign completed	Reset to 0
Promise / Refusal / Neutral	Next call +5 business days	Reset to 0
No answer (series < 5)	Next call +1 business day	+1 increment
No answer (series ≥ 5)	Campaign exhausted	+1 increment

CBS Integration and Offline Deployment

At 09:00 daily, a Celery task synchronizes the customer portfolio via an Oracle stored procedure (CBS). Deployment on the isolated network without Internet is accomplished by exporting Docker images as TAR archives, transferring via SCP, and loading on the target server. The procedure is automated with PowerShell scripts.

VII. DISCUSSION

Comparison with Commercial Platforms

TABLE XIII. COMPARISON WITH COMMERCAIL PLATFORMS

Feature	Our System	Commercial (Amazon Connect, Nuance)
Data sovereignty	Full on-premise	Cloud-dependent
Kyrgyz language	Yes (specialized models)	Not available
Isolated network	Supported	Requires Internet
Cost per call	Infrastructure only	Per-minute + API fees
Prompt flexibility	Full	Platform-constrained
Setup complexity	High	Low

Limitations and Improvement Paths

The primary limitation is the absence of telephony training data (8 kHz) for the Kyrgyz language. Mitigation paths include:

- Collection and labeling of real banking calls for fine-tuning the model on narrowband data;
- Domain adaptation through synthetic degradation of the existing corpus to 8 kHz;

- Streaming STT/TTS to reduce end-to-end latency below 2 seconds;
- Acoustic echo cancellation to support barge-in (interrupting the agent's speech).

Current end-to-end latency is 3.5–4.0 seconds, exceeding the 2-second target threshold. The primary contributor is LLM inference (~47% of total time). An LRU cache (200 entries) for repeated TTS phrases reduced the synthesis share for repeated sentences by ~35%.

VIII. CONCLUSION

This paper presents a production AI-based call center system demonstrating that open technologies — self-hosted LLM Qwen3, Whisper STT, and neural TTS — can be integrated into a traditional telephony production voice agent in a regulated banking environment with data sovereignty and air-gapped network requirements.

Empirical testing with 100 bank employees quantitatively confirmed: (1) high performance for Russian and English (WER 5–12%, completion rate 85–90%); (2) acceptable quality for Kyrgyz in short utterances with significant degradation in longer utterances; (3) effectiveness of the four-level error handling mechanism.

The Kyrgyz pipeline represents one of the first documented production deployments of a complete AI voice pipeline for the Kyrgyz language in banking. The key unresolved issue — absence of telephony training data (8 kHz) for Kyrgyz STT models — is a structural limitation affecting not only this system but any future voice application for Kyrgyz over telephony channels

Directions for future research: collection of a labeled corpus of Kyrgyz telephone speech (8 kHz), streaming STT/TTS, acoustic echo cancellation, and adaptation to other Central Asian markets (Kazakhstan, Uzbekistan, Tajikistan).

IX. REFERENCES

[1] R. Adamopoulou and L. Moussiades, “An overview of chatbot technology,” in IFIP Advances in Information and Communication Technology, vol. 584, pp. 373–383, 2020.

[2] M. McTear, “Conversational AI: Dialogue systems, conversational agents, and chatbots,” Synthesis Lectures on Human Language Technologies, 2020.

[3] J. Trivedi, “Examining the customer experience of using banking chatbots,” Journal of Internet Commerce, vol. 18, no. 1, pp. 91–111, 2019.

[4] G. Gupta et al., “Study of chatbot in customer service at banking,” in Proc. IC3I, 2023.

[5] A. Radford et al., “Robust speech recognition via large-scale weak supervision,” in Proc. ICML, 2023.

[6] K. Mussakhojayeva et al., “KazSpeech: A Kazakh speech corpus,” in Proc. Interspeech, 2023.

[7] A. Mukhamadiyev et al., “Development of language models for Uzbek speech recognition,” Sensors, 2023.

[8] T. Young et al., “Recent trends in deep learning based NLP,” IEEE Computational Intelligence Magazine, 2018.

[9] Qwen Team, “Qwen2.5 technical report,” 2024, arXiv:2412.15115. [Online]. Available: https://arxiv.org/abs/2412.15115

[10] H. Touvron et al., “LLaMA: Open and efficient foundation language models,” 2023, arXiv:2302.13971. [Online]. Available: https://arxiv.org/abs/2302.13971

[11] G. Skantze, “Turn-taking in conversational systems,” Computer Speech & Language, 2021.

[12] B. Spell, “hifi-gan-bwe: HiFi-GAN bandwidth extension,” GitHub, 2022. [Online]. Available: https://github.com/

[13] Silero Team, “Silero TTS: Quality benchmarks,” GitHub Wiki, 2023. [Online]. Available: https://github.com/

[14] S. Mehta et al., “MatchaTTS,” in Proc. ICASSP, 2024.

[15] J. Kong, J. Kim, and J. Bae, “HiFi-GAN: Generative adversarial networks for efficient and high fidelity speech synthesis,” in Proc. NeurIPS, 2020.

[16] A. Vaswani et al., “Attention is all you need,” in Proc. NeurIPS, 2017.

Farklı Kelime Gömme Yöntemlerinin CTI-NER Üzerine Etkisi

Utku Sobutay
Bilgisayar Mühendisliği Anabilim Dalı
Karabük Üniversitesi
Karabük, Türkiye
ORCID: 0000-0001-6700-7280

Oğuz Fındık
Bilgisayar Mühendisliği Anabilim Dalı
Karabük Üniversitesi
Karabük, Türkiye
ORCID: 0000-0001-5069-6470

***Özet*— Siber tehdit istihbaratı (Cyber Threat Intelligence - CTI), siber saldırıların erken tespiti, analizi ve önlenmesi açısından kritik bir rol oynamaktadır. Günümüzde artan siber tehdit hacmi ve karmaşıklığı, tehdit raporları, güvenlik bültenleri ve teknik analiz dokümanları gibi metin tabanlı kaynaklardan anlamlı bilgilerin otomatik olarak çıkarılmasını zorunlu hale getirmiştir. Bu noktada, Adlandırılmış Varlık Tanıma (Named Entity Recognition - NER) yöntemleri; tehdit aktörleri, zararlı yazılım adları, zafiyetler ve saldırı teknikleri gibi kritik varlıkların tespit edilmesinde temel bir araç olarak öne çıkmaktadır. Bununla birlikte, siber güvenlik alanına özgü terminoloji, düzensiz metin yapıları ve hızla değişen kavramlar, NER sistemlerinin başarımını doğrudan etkilemekte ve uygun kelime gömme (word embedding) yöntemlerinin seçimini kritik bir araştırma problemi haline getirmektedir.**

Bu çalışma, siber tehdit istihbaratı bağlamında farklı kelime gömme tekniklerinin CTI-NER performansı üzerindeki etkisini incelemeyi amaçlamaktadır. Bu amaç doğrultusunda, DNRTI veri seti kullanılarak Bag of Words (BoW), TF-IDF, Word2Vec, GloVe, FastText, BERT, RoBERTa, ALBERT, SecureBERT ve XLNet olmak üzere toplam on farklı kelime gömme yöntemi değerlendirilmiştir. Söz konusu gömme temsilleri, adlandırılmış varlık tanıma sürecinde yaygın olarak tercih edilen ve güçlü sonuçlar üreten iki temel derin öğrenme mimarisi BiLSTM ve BiLSTM-CRF ile entegre edilerek eğitim ve test aşamaları gerçekleştirilmiştir.

Elde edilen sonuçlar, bağlamsal ve transformer tabanlı yöntemlerin, siber tehdit istihbaratı verileri üzerinde klasik ve statik kelime gömme yaklaşımlarına göre daha yüksek performans sergilediğini ortaya koymaktadır. Özellikle, siber güvenlik alanına özgü olarak geliştirilmiş bağlamsal bir transformer modeli olan SecureBERT kelime gömme yönteminin BiLSTM-CRF ile birlikte kullanılması 0.834906 f1 skoru ile en yüksek başarıyı elde etmiştir.

***Keywords*— Adlandırılmış Varlık Tanıma, CTI-NER, Kelime Gömme Yöntemleri, Siber Güvenlik, Siber Tehdit İstihbaratı**

I. GIRIŞ

Günümüzün yoğun ağ bağlantılarına sahip hızla gelişen dijital ekosisteminde, siber saldırıların karmaşıklığı ve yıkıcılığı benzeri görülmemiş bir boyuta ulaşmıştır [1]. Gelişmiş Sürekli Tehditler (Advanced Persistent Threats - APT), fidye yazılımları ve sıfır gün (zero-day) zafiyetleri gibi evrilen tehditlere karşı koymak için kurumlar, reaktif savunma stratejilerinden proaktif bir yaklaşıma geçiş yapmak zorunda kalmışlardır [2]. Bu proaktif geçişin merkezinde, saldırganların taktik, teknik, prosedürlerini (Tactics, Techniques and Procedures - TTP) ve uzlaşma göstergelerini (Indicators of Compromise - IoC) analiz eden Siber Tehdit İstihbaratı (Cyber Threat Intelligence - CTI) yer almaktadır [3]. Ancak, tehdit istihbaratı verileri genellikle güvenlik "ölçüde yapılandırılmamış (unstructured) serbest metin formatında elde edilmektedir [1]. Bu devasa metin yığınları içerisindeki IP adresleri, zararlı yazılım aileleri, saldırgan grupları ve zafiyetler gibi kritik varlıkların insan gücüyle manuel olarak analiz edilmesi hem zaman alıcı hem de ölçeklenebilir olmamaktadır [4]. Bu darboğazı aşmak ve metinleri makine tarafından okunabilir yapısal bir formata dönüştürmek için Adlandırılmış Varlık Tanıma (Named Entity Recognition - NER) teknolojisi, CTI analizinin en kritik adımlarından biri haline gelmiştir [5].

NER sistemlerinin ve genel olarak makine öğrenmesi modellerinin doğal dil işleme (NLP) görevlerindeki başarısı, metinsel verilerin bilgisayarlar tarafından işlenebilecek sayısal vektörlere dönüştürülmesini sağlayan kelime gömme (word embedding) yöntemlerinin kalitesine doğrudan bağlı olmaktadır [6]. NLP literatüründe metin temsili, tarihsel olarak farklı evrelerden geçmiştir. Klasik yöntemler olan Bag of Words (BoW) ve TF-IDF, metinleri kelime frekanslarına veya nadirliklerine göre vektörlere çevirse de kelimelerin sözdizimsel sırasını ve anlamsal bağlamını modellemede yetersiz kalmaktadır [7]. Bu istatistiksel yaklaşımların kısıtlamalarını aşmak için geliştirilen Word2Vec, GloVe ve FastText gibi statik sinir ağı tabanlı gömme algoritmaları, kelimeleri anlamsal uzaylarda temsil ederek büyük bir ilerleme kaydetmiştir [8]. Özellikle FastText, alt-kelime (n-gram) düzeyinde çalışarak siber güvenlik metinlerinde sıkça karşılaşılan sözlük dışı (OOV) teknik terimlerin vektöre çevrilmesinde önemli bir avantaj sağlamaktadır [9]. Ancak statik yöntemlerin tümü, "shell", "patch" veya "virus" gibi aynı kelimenin farklı bağlamlarda kazandığı farklı anlamları (eşseslilik/çok anlamlılık) ayırt etme konusunda başarısız olmaktadır [2].

Derin öğrenme ve Transformer mimarilerindeki ilerlemeler, bağlamı dikkate alan gelişmiş dil modellerinin geliştirilmesini mümkün kılarak doğal dil işleme alanında köklü bir dönüşüm yaratmıştır [10]. BERT, RoBERTa ve ALBERT gibi önceden eğitilmiş devasa dil modelleri, cümlenin çift yönlü bağlamını analiz ederek kelimelerin cümle içindeki pozisyonlarına göre dinamik vektörler üretebilmektedir [11]. XLNet ise; permütasyon tabanlı dil modellemesi kullanarak, BERTin maskeleme kısıtlamalarını aşmakta ve uzun CTI raporlarında birbirine uzak konumdaki siber varlıklar arasındaki uzun menzilli bağımlılıkları (long-range dependencies) etkili bir şekilde yakalayabilmektedir [5]. Bununla birlikte, genel amaçlı metinler üzerinde eğitilen dil modelleri, siber güvenlik alanındaki özel terminoloji ve alan-spesifik ifadelerin semantik yapısını yeterli düzeyde öğrenememektedir [12]. Bu anlamsal boşluğu doldurmak amacıyla, doğrudan siber tehdit istihbaratı metinleriyle ön eğitime tabi tutulan SecureBERT gibi alan-özgü (domain-specific) dil modelleri geliştirilmiştir [12].

Kelime gömme yöntemleriyle oluşturulan bağlamsal vektörlerin, siber güvenlik metinlerindeki uzun ve karmaşık

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

varlık sınırlarını doğru tespit edebilmek amacıyla karar mekanizmalarıyla birlikte kullanılması gerekmektedir. Çift Yönlü Uzun Kısa Süreli Bellek (Bidirectional Long Short Term Memory - BiLSTM) ağları, metni hem ileri hem de geri yönlü işleyerek ardışık ve zamansal bağımlılıkları yakalarken, Koşullu Rastgele Alanlar (Conditional Random Fields - CRF) katmanı, etiketler arası geçiş olasılıklarını (transition probabilities) modelleyerek geçersiz etiket dizilimlerini engeller ve küresel olarak en uygun diziyi tahmin edebilmektedir [1].

Literatürde farklı kelime gömme modellerinin veya derin öğrenme tabanlı mimarilerin siber güvenlik metinleri üzerinde tekil olarak incelendiği çalışmalar bulunsa da, klasik yaklaşımlardan, en gelişmiş alana-özgü transformerlara kadar uzanan geniş bir yelpazenin, standart karar mekanizmaları üzerinde sistematik olarak karşılaştırıldığı araştırmalar oldukça sınırlıdır. Bu eksikliği gidermek amacıyla bu çalışmada, siber güvenlik alanında geniş çaplı ve referans kabul edilen DNRTI (Dataset for Named Entity Recognition in Threat Intelligence) veriseti kullanılarak kapsamlı bir analiz sunulmaktadır [5]. Çalışma kapsamında, yapılandırılmamış CTI raporlarından varlık çıkarımı görevi için 10 farklı kelime gömme yöntemi (Klasik Yöntemler: Bag of Words, TF-IDF; Statik Yöntemler: Word2Vec, GloVe, FastText; Transformer Tabanlı Yöntemler: BERT, RoBERTa, ALBERT, SecureBERT, XLNet) seçilmiştir. Seçilen kelime gömme yöntemlerinin siber güvenlik alanındaki varlık tanıma performansını değerlendirmek amacıyla, modeller hem temel BiLSTM hem de BiLSTM-CRF karar mekanizmaları üzerinde test edilmiştir.

Bu makale, farklı nesil metin temsil teknolojilerinin ve dizi etiketleme mimarilerinin CTI- NER performansına olan etkilerini karşılaştırmalı olarak ortaya koymayı ve operasyonel tehdit istihbaratı sistemleri için en verimli ardışık düzeni (pipeline) belirlemeyi hedeflemektedir.

II. Literatür Taraması

A. Klasik Yöntemler (Bag of Words ve TF-IDF)

Literatürde yapılan karşılaştırmalı çalışmalarda, Bag of Words ve TF-IDF gibi klasik yöntemlerin, Destek Vektör Makineleri (SVM), Rastgele Orman (Random Forest) ve Lojistik Regresyon gibi geleneksel makine öğrenmesi algoritmalarıyla birleştirildiğinde metin sınıflandırmada (0.93 doğrulukla) kabul edilebilir sonuçlar verebildiği görülmüştür [13]. Ancak siber güvenlik NER görevlerinde, kelimelerin sözdizimsel sırasını ve anlamsal bağlamını tamamen göz ardı ettiklerinden dolayı bu yöntemler, karmaşık tehdit istihbaratı metinlerinden bilgi çıkarımı için yetersiz kalmaktadır [7].

Shin, Kwon ve Ryu çalışmalarında, Twitter tabanlı CTI tespitinde TF-IDF ve benzeri geleneksel özellik çıkarımı yöntemlerini, Word2Vec, GloVe ve FastText gibi gömme yöntemleriyle karşılaştırmış; TF-IDF'nin açıklanabilir ve hızlı olmasına rağmen, bağlamı ve alan terimlerinin anlam ilişkilerini yakalamada sınırlı kaldığını göstermiştir [14].

Haq, Khan ve Alshehri ise; tehdit tespitinde TF-IDF ile Word2Vec ve GloVe temsillerini kullanarak klasik temsillerin güvenlik metinlerinde başlangıç seviyesi için yararlı olduğunu, ancak anlamsal yakınlık gerektiren çıkarım görevlerinde statik gömmelerin daha güçlü sonuç verdiğini belirtmiştir [15].

B. Statik Kelime Gömme Yöntemleri (Word2Vec, GloVe, FastText)

Statik kelime gömme yöntemleri olan Word2Vec, GloVe ve FastText, siber güvenlik metinlerinde terimler arası anlamsal benzerliği öğrenebilmesi nedeniyle klasik yöntemlere göre daha güçlü bir temsil sunmaktadır. Bununla beraber, Word2Vec ve GloVe yöntemleri kelimeler arası anlamsal yakınlığı yakalamada başarılı olsalar da siber güvenlikte sıkça karşılaşılan sözlük dışı (Out of Vocabulary - OOV) karmaşık dosya yolları, IP adresleri ve yeni zararlı yazılım isimlerini temsil etmekte zorlanmaktadır. Özellikle CNN mimarisi ile GloVe kelime gömmelerinin entegre edildiği çalışmalarda genellikle 0.84-0.85 bandında doğruluk elde edilebilmektedir [13].

FastText kelime gömme yönteminin, kelimeleri alt-karakter tabanlı n-gram yapılarına ayırabilmesi sayesinde siber güvenlik alanındaki bilinmeyen (OOV) veya nadir terimleri modellemede Word2Vec ve GloVe yöntemlerine kıyasla daha başarılı sonuçlar verdiği çeşitli LSTM ve CNN tabanlı çalışmalarda gösterilmiştir [13].

C. Transformer Tabanlı Yöntemler (BERT, RoBERTa, ALBERT, XLNet)

Transformer tabanlı yöntemler, cümle bağlamını çift yönlü olarak analiz edebilme yetenekleri sayesinde eşsesli kelimelerden kaynaklanan çok anlamlılık (polysemy) problemini daha etkin biçimde çözebilmekte ve DNRTI gibi CTI veri setleri üzerindeki CTI-NER görevlerinde üstün performans sergilemektedir.

Chen ve arkadaşları tarafından DNRTI veriseti üzerinde yapılan çalışmalarında, genel İngilizce ile eğitilmiş BERT modelinin doğrudan CRF ile birleştirildiği BERT-CRF modelinin 0.9002 gibi oldukça yüksek bir F1 skoruna ulaştığı, BERT-BiLSTM-CRF modelinin ise DNRTI üzerinde 0.8459 F1 skoru elde ettiği raporlanmıştır [4].

Evangelatos ve arkadaşları çalışmalarında, DNRTI veri setindeki 300'den fazla açık kaynak tehdit istihbaratı raporu üzerinde BERT, RoBERTa ve ALBERT gibi transformer modellerini incelemiş ve bu modellerin önceki yöntemlere göre siber güvenlik varlıklarını çıkarmada belirgin biçimde daha etkili olduğunu göstermiştir [16].

RoBERTa modeli, BERT modelinin optimize edilmiş bir veriyonu olarak daha uzun metinlerde daha başarılı sonuçlar sağlarken, ALBERT modeli, parametre paylaşımı ile donanım maliyetini düşürerek uç cihazlarda çalışabilen siber tespit sistemleri için verimli bir alternatif oluşturmaktadır [10].

Wang, Liu ve arkadaşları çalışmalarında, birbirinden uzak siber varlıklar arasındaki uzun menzilli bağımlılıkları çözmek için permütasyon tabanlı dil modellemesi yapan XLNet-CRF modelini tanıtmışlardır. Bu model, CTI-Reports veri setinde 0.9743 F1 skoru ile üstün bir başarı yakalamıştır. Fakat, DNRTI veri setinin kısa, e-posta formatına benzeyen ("admin@338") ve yoğun etiketli yapısında BERT modelinin maskeleme mimarisinin daha avantajlı olması sebebiyle, XLNet-CRF modeli DNRTI üzerinde 0.8239 F1 skorunda kalmış; aynı veri setinde BERT-CRF modeli XLNet modelini geçerek 0.9002 F1 skoru elde etmiştir [5].

D. Alana Özgü Transformer Tabanlı Yöntemler (SecureBERT, SecureBERT 2.0)

Doğrudan siber tehdit istihbaratı verileri üzerinde ön eğitime tabi tutulan alana özgü transformer modelleri, genel

amaçlı dil modellerinin alan bilgisindeki eksikliklerini azaltarak, siber güvenlik uygulamalarında temel yaklaşım olarak tercih edilir hale gelmiştir. Özellikle, Aghaei ve arkadaşları tarafından geliştirilen SecureBERT modeli, siber güvenlik metinleriyle eğitilmiş ve NER, metin sınıflandırma gibi görevlerde geleneksel modelleri geride bırakmıştır [12]. Bununla birlikte, modern bir mimari ile 13 milyar metin ve 53 milyon kod parçacığından (token) oluşan geniş ölçekli siber güvenlik veri kümeleri ile eğitilen SecureBERT 2.0 modeli, siber güvenlik NER testlerinde 0.945 F1 skoru elde ederek dikkat çekici bir başarı göstermiştir [17].

Li ve arkadaşları, DNRTI ve MalwareDB verisetleri kullanarak siber güvenlik bilgi grafiği oluşturmak amacıyla SecureBERT_Plus, BiLSTM, Attention ve CRF katmanlarına sahip bir model önermiştir ve önerilen bu model, DNRTI üzerinde 0.839 genel F1 skoru ve örtüşen ilişkilerde yüksek başarı göstererek karmaşık saldırı zincirlerini başarıyla yapılandırmıştır [1].

Chen ve arkadaşları, DNRTI üzerinde uzun ve çoklu kelimeden oluşan varlıkların çıkarımını iyileştirmek için Segment Maskeleme Geliştirmesi (SME) yöntemini içeren SecureBERT_SME modelini eğitmişlerdir. SSNER ismini verdikleri bu model, DNRTI üzerinde 0.8612 F1 skoru elde etmiştir [2].

III. VERISETI

Bu çalışmada, farklı kelime gömme yöntemlerinin CTI-NER üzerindeki performansını analiz etmek amacıyla, siber tehdit istihbaratı alanında referans niteliği taşıyan DNRTI (Dataset for Named Entity Recognition in Threat Intelligence) verisetinden yararlanılmıştır. Veriseti, %70 eğitim, %30 test ve %50 doğrulama olmak üzere ayrılmıştır.

DNRTI veriseti, CTI metinlerinden yapılandırılmış bilgi çıkarmak ve NER modellerini eğitmek amacıyla geliştirilmiş büyük ölçekli ve referans kabul edilen bir verisetidir [18]. Geleneksel NER verisetleri siber güvenlik terminolojisinin karmaşık yapısını temsil etmede yetersiz kaldığından dolayı, Wang ve arkadaşları [19] tarafından bu alandaki eksikliği gidermek üzere özel olarak tasarlanmıştır.

DNRTI, internet üzerinden toplanan 300'den fazla gerçek dünya siber tehdit raporunun analiz edilmesiyle oluşturulmuş 6576 cümle ve BIO (Begin, Inside, Outside) formatında işaretlenmiş yaklaşık 182452 kelime içermektedir [4].

Verisetinde manuel olarak etiketlenmiş siber tehditleri oluştruan kavramlardan oluşan 13 farklı kategoride 36.812 adet varlık (entity) bulunmaktadır [3]. Bu varlık kategorisinin veriseti içerisindeki sayıları ve genel dağılım oranları Tablo I' de sunulmuştur [5].

TABLO I. DNRTI VARLIK KATEGORILERI, SAYILARI VE ORANLARI

Kategori Etiketi	*Kategori Açıklaması*	*Varlık Sayısı*	*Yüzde (%)*
HackOrg	Hacker Organizasyonu	5465	14.8
Tool	Saldırı Araçları	4784	13.0
Area	Kaynak veya Hedef Konum	3447	9.4
OffAct	Saldırı Eylemi / Aksiyonu	2669	7.3
Time	Olayın Gerçekleştiriği Tarih / Saat	2659	7.2
Org	Hedef Organizasyon	2489	6.8
Features	Saldırının Teknik Nitelikleri	2441	6.6
Purp	Amaç	2424	6.6
SamFile	Dosya İsimleri / Hash Değerleri	2400	6.5
Idus	Hedef Alınan Endüstri / Sektör	2136	5.8
Way	Yöntem / Teknik	2018	5.5
Exp	Zafiyet / İstismar	1959	5.3
SecTeam	Analiz Eden Güvenlik Ekibi	1921	5.2

IV. METODOLOJI

Çalışmada, farklı kelime gömme yöntemlerinin CTI-NER performansı üzerindeki etkilerini değerlendirmek amacıyla; Bag of Words ve TF-IDF gibi klasik yaklaşımlar, Word2Vec, GloVe ve FastText gibi statik kelime gömme yöntemleri ve BERT, RoBERTa, ALBERT, SecureBERT ve XLNet gibi transformer tabanlı modeller, BiLSTM ve BiLSTM-CRF mimarileri üzerinde ayrı ayrı test edilmiştir.

A. Klasik Yöntemler

Bu yöntemler, kelimeleri sayısal vektörler ile temsil edebilmek amacıyla kelime gömme yöntemleri öncesinde dönemlerde kullanılan istatistiksel yöntemlerdir [20].

Bu çalışmada, klasik yöntemler için maksimum öznitelik (max features) sayısı 10000 olarak belirlenmiştir.

- *Bag of Words (BoW):* Metinler içerisinde yer alan tüm benzersiz kelimelerden bir sözlük oluşturmakta ve her metin, bu kelimelerin geçiş sıklıklarına göre bir vektör olarak ifade edilmektedir. Uygulanması ve anlaşılması basittir, eğitim gerektirmez ve doğrudan sayım yaptığı için oldukça hızlı çalışmaktadır. Fakat bu yöntem kelimelerin sözdizimsel sırasını, bağlamını ve anlamsal özelliklerini tamamen göz ardı etmektedir [7].
- *TF-IDF*: Bu yöntem, BoW modelinin her kelimeyi eşit derecede önemli sayması zafiyetini gidermek amacıyla geliştirilmiş, kelimelerin "nadirliğine" odaklanan bir yöntemdir [20]. Metinlerde çok sık geçen ancak ayırt edici bir bilgi taşımayan yaygın kelimelerin ağırlığını otomatik olarak düşürmekte, buna karşın, dokümana özgü olan nadir ve anlamlı kelimelerin önemini artırmaktadır. Fakat, BoW gibi TF-IDF de kelimenin bağlamını, sözdizimi sırasını ve kelimeler arasındaki anlamsal ilişkileri modelleyememektedir [7].

B. Statik Kelime Gömme Yöntemleri

Bu yöntemler, kelimeler arasındaki anlamsal yakınlıkları dikkate alarak, kelimeleri düşük boyutlu vektörler olarak temsil eden yöntemlerdir [8]. Klasik istatiksel yöntemlerin aksine temelinde yapay sinir ağları kullanılmaktadır [21].

Statik kelime gömme yöntemlerinde her bir kelime için oluşturulacak vektör boyutları önceden belirlenebilmektedir [21]. Bu çalışmada, her bir kelime gömme yöntemi için 50, 100, 200 ve 300 vektör boyutları test edilmiştir. Pencere boyutu (window size) olarak 2 tercih seçilmiştir.

- *Word2Vec*: 2013 yılında Tomas Mikolov ve ekibi tarafından geliştirilen Word2Vec, kelimeleri yerel bağlamlarına göre vektör uzayında temsil etmektedir [21]. Bu yöntemde iki temel sinir ağı mimarisi kullanılmaktadır: Continuous Bag of Words (CBOW) mimarisinde; herhangi bir kelimenin çevresindeki

bağlam kelimelerinden yola çıkarak merkezdeki hedef kelimeyi tahmin etmeye, Skip-Gram mimarisinde ise; verilen bir merkez kelimesinden yola çıkarak, kelime etrafındaki bağlam kelimelerini tahmin etmeye çalışmaktadır [9]. Word2Vec, dilbilimsel örüntüleri doğrusal ilişkiler olarak öğrenmede başarılıdır, ancak kelimeleri bölünemez bir bütün olarak ele aldığından dolayı eğitim setinde yer almayan sözlük dışı kelimelere karşılık gelen bir vektör üretememektedir [22]. Çalışmada; "Gensim Word2Vec" kullanılmıştır [23] ve CBOW modeli tercih edilmiştir.

- *GloVe*: Pennington ve arkadaşları tarafından 2014 yılında Stanford Üniversitesi'nde geliştirilen GloVe modeli, Word2Vec' in yalnızca yerel bağlam pencerelerine dayalı yaklaşımını, tüm verisetinin küresel eş-oluşum istatistikleri ve matris çarpanlarına ayırma teknikleriyle birleştirmektedir [24]. Hedef, iki kelimenin vektörlerinin iç çarpımının, bu iki kelimenin birlikte geçme sıklığının logaritmasına eşit olmasını sağlamaktır [9]. Ancak GloVe da tıpkı Word2Vec gibi sözlük dışı (OOV) kelimeleri temsil etmekte başarısız olmaktadır [22].

- *FastText*: Facebook yapay zeka araştırma ekibi bünyesinde bulunan Bojanowski ve arkadaşları tarafından Word2Vec' in Skip-gram mimarisinin bir uzantısı olarak 2017 yılında geliştirilen bu yöntem, kelimeleri tek bir birim olarak değil, alt-karakter dizilerinin (n-gram) toplamı olarak ele almaktadır [25]. FastText'in en büyük avantajı morfolojik analiz yeteneğidir ve eğitim sırasında hiç görülmemiş nadir veya yeni kelimelerin vektörleri ile o kelimeyi oluşturan karakter n-gramlarının vektörleri toplanarak elde edilebilmektedir [22]. Bu esneklik, özellikle sondan eklemeli dillerin işlenmesinde ve siber güvenlikte sıkça karşılaşılan yeni nesil zararlı yazılım isimleri, IP adresleri ve dosya yolları gibi standart dışı teknik kelimelerin vektör olarak temsil edilmesinde FastText yöntemini, Word2Vec ve GloVe' dan üstün kılmaktadır [8]. Bu çalışmada; "Gensim FastText" kullanılmıştır [26] ve CBOW modeli tercih edilmiştir.

C. Transformer Tabanlı Yöntemler

Word2Vec, GloVe ve FastText gibi statik kelime gömme yöntemlerinin temel sınırlılıklarından biri, aynı kelimenin farklı bağlamlardaki anlamsal değişimlerini yeterince temsil edememeleridir. Buna karşılık transformer mimarileri, derin öğrenme mimarileri ve dikkat (attention) mekanizmaları aracılığıyla cümleyi çift yönlü analiz ederek bağlama duyarlı dinamik kelime temsilleri üretmektedir [10].

Transformer tabanlı yöntemlerde kullanılan maksimum dizi uzunluğu (maximum length), modele tek seferde verilebilecek en fazla kelime/parça sayısını ifade etmektedir [27]. Bu çalışmada maksimum kelime/parça uzunluğu 256 olarak alınmıştır.

- *BERT*: Google araştırmacıları Devlin ve arkadaşları tarafından 2018 yılında geliştirilen BERT modeli, Maskelenmiş Dil Modellemesi (Masked Language Modeling - MLM) yaklaşımı sayesinde metni çift yönlü olarak analiz ederek kelimenin bağlamsal anlamı öğrenebilmektedir [27]. Cümledeki kelimelerin %15' ini rastgele maskeleyip bu maskelenmiş kelimeleri tahmin etmeye çalışarak eğitilmektedir [12]. Bu model, siber güvenlik metinlerindeki anlamsal ilişkileri ve bağlamı yakalamada statik yöntemlerden çok daha başarılı olmasına rağmen genel İngilizce metinler kullanılarak eğitilmesinden dolayı siber güvenlik terminolojisini tam anlamıyla temsil edememektedir [5].

- *RoBERTa*: Facebook AI araştırmacılarından Liu ve arkadaşları tarafından 2019 yılında geliştirilen bu model, BERT modelinin parametrelerinin optimize edilmiş bir versiyonu olarak sunulmuştur [28]. BERT modelinde bulunan "Bir Sonraki Cümle Tahmini" (Next Sentence Prediction – NSP) görevi kaldırılarak, daha büyük verisetleriyle ve "dinamik maskeleme" stratejileri kullanılarak çok daha uzun süre eğitilmiştir [10]. Bu iyileştirmeler sayesinde, karmaşık ve uzun siber tehdit raporlarının analizinde standart BERT modeline kıyasla daha yüksek bir metin anlama ve varlık tanıma performansı sunmaktadır [29].

- *ALBERT*: Siber güvenlik operasyon merkezlerinde veya donanım kısıtı olan sistemlerde ağır modellerin çalıştırılması maliyetli ve yavaş olmaktadır. Bu model 2020 yılında Lan ve arkadaşları tarafından BERT modelinin donanım gereksinimlerini ve hesaplama maliyetini düşürmek amacıyla tasarlanmıştır [30]. Gömme matrislerinin ayrıştırılması ve katmanlar arası parametre paylaşımı teknikleri kullanılarak model boyutu küçültürmüştür. Ayrıca NSP görevi yerine, metnin mantıksal akışını daha iyi kavramasını sağlayan "Cümle Sırası Tahmini" (Sentence Order Prediction - SOP) görevi eklenmiştir [10]. Bu model çok daha az parametre ile BERT modeline yakın bir başarı sunarak yüksek hızlı çıkarımlar yapma imkanı sağlamaktadır.

- *SecureBERT*: Aghaei ve arkadaşları tarafından 2022 yılında geliştirilen alana özgü bir transformer modelidir ve genel amaçlı modellerin siber güvenlik ile ilgili metinlerde hata yapmasını önlemek amacıyla doğrudan siber tehdit istihbaratı raporları, zafiyet veri tabanları ve teknik makaleler kullanılarak eğitilmiştir [12]. BERT modeline göre siber güvenlik ile ilgili görevlerde daha başarılıdır [17].

- *XLNet*: Bu model, BERT modelinin MLM yaklaşımının maskeli kelimelerinden kaynaklanan bağımsızlık varsayımı problemlerini ve uzun metinlerde ortaya çıkan bağlamsal bilgi kaybını gidermek amacıyla Yang ve arkadaşları tarafından 2019 yılında geliştirilmiştir [31]. XLNet, maskeleme tekniği yerine permütasyon tabanlı dil modellemesi kullanarak, bir cümledeki kelimelerin tüm olası sıralamalarını değerlendirmekte ve iki akışlı öz-dikkat mekanizması ile hedef kelimenin tahminini kendi içeriğinden bağımsız hale getirerek sızıntıyı önlemektedir [5]. CTI raporları genellikle uzun olmaktadır ve bir saldırı vektörünün başlangıcı ile hedef sistemin taviz vermesi arasında uzun cümleler bulunabilmektedir. XLNet, bu tarz birbirine uzak, kesintili ve iç içe geçmiş siber varlıklar arasındaki uzun menzilli bağımlılıkları modellemede geleneksel BERT mimarilerine karşı ciddi bir mimari üstünlük sunmaktadır [5].

D. Adlandırılmış Varlık Tanıma (NER)

Siber Tehdit İstihbaratı alanında Adlandırılmış Varlık Tanıma (NER), yapılandırılmamış CTI metinlerden zararlı yazılım (malware), zafiyet (vulnerability) ve tehdit aktörü gibi önceden tanımlanmış varlıkları otomatik olarak tespit edip sınıflandırmayı amaçlayan temel bir dizi etiketleme (sequence labeling) problemidir [32].

NER sistemlerinin geliştirilmesinde farklı istatistiksel ve makine öğrenmesi yöntemleri kullanılsana rağmen, özellikle karmaşık metinlerde bağlamsal ilişkileri ve yapısal tutarlılığı korumak konusunda başarılı derin öğrenme mimarilerinden BiLSTM ve BiLSTM-CRF modelleri yaygın olarak tercih edilmektedir [33].

Bu çalışma kapsamında; kelime gömme yöntemleri ile birlikte BiLSTM ve BiLSTM-CRF modelleri test edilmiştir. Modellerde, gizli katman boyutu (hidden layer) 256 ve eğitim döngü sayısı (epoch) 32 olarak belirlenmiştir. Klasik ve Statik yöntemlerde yığın boyutu (batch size) 32 ve öğrenme oranı (learning rate) 0,001, transformer tabanlı yöntemlerde yığın boyutu 8 ve öğrenme oranı 0.00002 olarak seçilmiştir.

- *BiLSTM*: Standart Tekrarlayan Sinir Ağları (RNN), uzun metin dizilerini işlerken "kaybolan gradyan" (vanishing gradient) probleminden etkilenmektedir; LSTM (Long Short-Term Memory) mekanizmaları bu sorunu çözmek için tasarlanmıştır [33]. BiLSTM modeli ise; bu yapıyı daha da ileri taşıyarak metin dizilerini sadece soldan sağa değil, aynı zamanda sağdan sola olmak üzere iki yönlü işlemektedir. Bu çift yönlü işleme kapasitesi, modelin bir kelimenin hem öncesindeki hem de sonrasındaki bağlam bilgisini eşzamanlı olarak yakalamasını ve kelime için zengin bir anlamsal vektör oluşmasını sağlamaktadır. Fakat bu model, her kelimenin etiketine ayrı ayrı ve bağımsız olarak karar verme eğilimindedir, bu durum dizisel kuralların ihlal edilmesine yol açabilmektedir [34].
- *CRF*: BiLSTM katmanının bağımsız tahmin yapma zayıflığını gidermek için modelin çıktı aşamasına CRF katmanı eklenmektedir [34]. CRF katmanı, etiketler arası geçiş olasılıklarını bir matris üzerinden modelleyerek geçersiz dizilimleri engellemektedir [35].
- *BiLSTM-CRF*: BiLSTM modeline entegre edilen CRF katmanı sayesinde, kelime düzeyinde elde edilen bağlamsal temsiller, dizisel etiket bağımlılıklarıyla birlikte değerlendirilerek daha tutarlı sonuçlar üretilmektedir [33]. Özellikle karmaşık yapıya sahip siber tehdit raporlarında ve uzun metinlerde varlık sınırlarının doğru çizilebilmesi, iç içe geçmiş olan varlıkların ayrıştırılabilmesi ve birbirine uzak olan bağlamların yakalanması için BiLSTM-CRF modeli günümüzde standart ve son derece güçlü bir çözüm olarak kabul edilmektedir [35].

E. Deneysel Ortam

Bu çalışmada, AMD Ryzen 7 1700x 3.40 Ghz işlemci, 32 GB 3200 Mhz DDR4 bellek, MS Windows 11 x64 işletim sistemi ve NVIDIA RTX 4070 ekran kartı donanım olarak kullanılmıştır. Yazılım ortamında ise; Python version 3.10.20 ve PyTorch kütüphanesi tercih edilmiştir.

Transformer tabanlı yöntemlerden BERT modeli için “bert-base-cased”, RoBERTa modeli için “roberta-base”, ALBERT modeli için “albert-base-v2”, SecureBERT modeli için “ehsanaghaei/SecureBERT” ve XLNet modeli için de “xlnet/xlnet-base-cased” önceden eğitilmiş modelleri test edilmiştir.

Çalışmada bahsedilen modellerin farklı kelime gömme yöntemleri ile karşılaştırılmasını yapabilmek amacıyla yaygın olarak kullanılan nicel değerlendirme ölçütlerinden biri olan F1 skoru temel alınmıştır. Bununla birlikte, her model için doğruluk (accuracy), kesinlik (presicion) ve geri çağırma (recall) değerleri de tablolar halinde sunulmuştur.

V. BULGULAR

Bu bölümde, 10 farklı kelime gömme yönteminin BiLSTM ve BiLSTM-CRF modelleri ile test sonuçları ayrı ayrı ele alınmıştır.

A. Klasik Yöntemler

Klasik yöntemler için BiLSTM modeli ile sonuçları Tablo II’de, BiLSTM-CRF modeli ile sonuçları Tablo III’te ve genel sonuçlar Tablo IV’te verilmiştir.

TABLO II. BILSTM SONUÇLARI

Model	Precision	Recall	F1	Accuracy
TFIDF-BiLSTM	0.725712	0.727572	0.726641	0.921406
BoW-BiLSTM	0.719740	0.690223	0.704673	0.915242

TABLO III. BILSTM-CRF SONUÇLARI

Model	Precision	Recall	F1	Accuracy
TFIDF-BiLSTM-CRF	0.767362	0.740388	0.753634	0.925482
BoW-BiLSTM-CRF	0.724415	0.692054	0.707865	0.913800

TABLO IV. GENEL SONUÇLAR

Model	Precision	Recall	F1	Accuracy
TFIDF-BiLSTM-CRF	0.767362	0.740388	0.753634	0.925482
TFIDF-BiLSTM	0.725712	0.727572	0.726641	0.921406
BoW-BiLSTM-CRF	0.724415	0.692054	0.707865	0.913800
BoW-BiLSTM	0.719740	0.690223	0.704673	0.915242

Tablo II ve Tablo II’e göre TF-IDF tabanlı modellerin BoW tabanlı modellere göre daha başarılı sonuçlar ürettiği dikkat çekmektedir. Bunun temel nedeni TF-IDF yönteminin yalnızca kelime frekansını değil, aynı zamanda kelimenin veri kümesi içerisindeki ayırt ediciliğini de dikkate almasıdır [7]. Böylece siber güvenlik metinlerinde sık geçen ancak düşük bilgi değeri taşıyan kelimelerin etkisi azaltılarak daha anlamlı özellik temsilleri elde edilmektedir.

Tablo IV’te TF-IDF-BiLSTM-CRF modelinin 0.753634 F1 skorunun diğer tüm klasik yöntemlerden yüksek olması, TF-IDF ile CRF katmanının birlikte kullanılmasının varlık sınırlarını daha tutarlı şekilde öğrenebildiğini göstermektedir. CRF katmanı, BIO etiketleri arasındaki geçiş ilişkilerini modellediği için özellikle çok sözcüklü veya karmaşık siber güvenlik varlıklarının tanınmasında avantaj sağlamaktadır [35].

B. Statik Kelime Gömme Yöntemleri

Statik kelime gömme yöntemlerinin 50, 100, 200 ve 300 vektör boyutundaki BiLSTM modeli ile sonuçları sırasıyla Tablo V, Tablo VII, Tablo IX ve Tablo XI'de ve BiLSTM-CRF modeli ile sonuçları Tablo VI, Tablo VIII, Tablo X ve Tablo XII'de, genel sonuçlar ise; Tablo XIII'te sunulmuştur.

TABLO V. 50 VEKTÖR BOYUTUNDA BILSTM SONUÇLARI

Model	Precision	Recall	F1	Accuracy
Word2Vec-BiLSTM (50)	0.729093	0.759795	0.744128	0.927073
GloVe-BiLSTM (50)	0.738456	0.749542	0.743958	0.926576
FastText-BiLSTM (50)	0.714870	0.767484	0.740244	0.924886

TABLO VI. 50 VEKTÖR BOYUTUNDA BILSTM-CRF SONUÇLARI

Model	Precision	Recall	F1	Accuracy
GloVe-BiLSTM-CRF (50)	0.784998	0.774075	0.779499	0.932044
Word2Vec-BiLSTM-CRF (50)	0.765720	0.780300	0.772942	0.929857
FastText-BiLSTM-CRF (50)	0.760073	0.766752	0.763398	0.927968

TABLO VII. 100 VEKTÖR BOYUTUNDA BILSTM SONUÇLARI

Model	Precision	Recall	F1	Accuracy
GloVe-BiLSTM (100)	0.758278	0.754669	0.756469	0.929708
FastText-BiLSTM (100)	0.740871	0.772611	0.756408	0.929608
Word2Vec-BiLSTM (100)	0.747486	0.761992	0.754669	0.929658

TABLO VIII. 100 VEKTÖR BOYUTUNDA BILSTM-CRF SONUÇLARI

Model	Precision	Recall	F1	Accuracy
Word2Vec-BiLSTM-CRF (100)	0.793734	0.779202	0.786401	0.934778
GloVe-BiLSTM-CRF (100)	0.790175	0.759795	0.774687	0.931845
FastText-BiLSTM-CRF (100)	0.740663	0.777005	0.758399	0.925234

TABLO IX. 200 VEKTÖR BOYUTUNDA BILSTM SONUÇLARI

Model	Precision	Recall	F1	Accuracy
GloVe-BiLSTM (200)	0.781618	0.753570	0.767338	0.931298
FastText-BiLSTM (200)	0.770310	0.763823	0.767053	0.932342
Word2Vec-BiLSTM (200)	0.749911	0.774075	0.761802	0.930603

TABLO X. 200 VEKTÖR BOYUTUNDA BILSTM-CRF SONUÇLARI

Model	Precision	Recall	F1	Accuracy
Word2Vec-BiLSTM-CRF (200)	0.783543	0.774075	0.778781	0.931597
GloVe-BiLSTM-CRF (200)	0.794039	0.760893	0.777113	0.929658
FastText-BiLSTM-CRF (200)	0.759520	0.788722	0.773846	0.929111

TABLO XI. 300 VEKTÖR BOYUTUNDA BILSTM SONUÇLARI

Model	Precision	Recall	F1	Accuracy
Word2Vec-BiLSTM (300)	0.753965	0.783230	0.768319	0.932094
GloVe-BiLSTM (300)	0.747624	0.777737	0.762383	0.929459
FastText-BiLSTM (300)	0.745353	0.778103	0.761376	0.930006

TABLO XII. 300 VEKTÖR BOYUTUNDA BILSTM-CRF SONUÇLARI

Model	Precision	Recall	F1	Accuracy
GloVe-BiLSTM-CRF (300)	0.808812	0.752838	0.779822	0.931845
Word2Vec-BiLSTM-CRF (300)	0.788563	0.767484	0.777881	0.932591
FastText-BiLSTM-CRF (300)	0.756947	0.787990	0.772156	0.926974

TABLO XIII. GENEL SONUÇLAR

Model	Precision	Recall	F1	Accuracy
Word2Vec-BiLSTM-CRF (100)	0.793734	0.779202	0.786401	0.934778
GloVe-BiLSTM-CRF (300)	0.808812	0.752838	0.779822	0.931845
GloVe-BiLSTM-CRF (50)	0.784998	0.774075	0.779499	0.932044
Word2Vec-BiLSTM-CRF (200)	0.783543	0.774075	0.778781	0.931597
Word2Vec-BiLSTM-CRF (300)	0.788563	0.767484	0.777881	0.932591
GloVe-BiLSTM-CRF (200)	0.794039	0.760893	0.777113	0.929658
GloVe-BiLSTM-CRF (100)	0.790175	0.759795	0.774687	0.931845
FastText-BiLSTM-CRF (200)	0.759520	0.788722	0.773846	0.929111
Word2Vec-BiLSTM-CRF (50)	0.765720	0.780300	0.772942	0.929857
FastText-BiLSTM-CRF (300)	0.756947	0.787990	0.772156	0.926974
Word2Vec-BiLSTM (300)	0.753965	0.783230	0.768319	0.932094
GloVe-BiLSTM (200)	0.781618	0.753570	0.767338	0.931298
FastText-BiLSTM (200)	0.770310	0.763823	0.767053	0.932342
FastText-BiLSTM-CRF (50)	0.760073	0.766752	0.763398	0.927968
GloVe-BiLSTM (300)	0.747624	0.777737	0.762383	0.929459
Word2Vec-BiLSTM (200)	0.749911	0.774075	0.761802	0.930603
FastText-BiLSTM (300)	0.745353	0.778103	0.761376	0.930006
FastText-BiLSTM-CRF (100)	0.740663	0.777005	0.758399	0.925234
GloVe-BiLSTM (100)	0.758278	0.754669	0.756469	0.929708
FastText-BiLSTM (100)	0.740871	0.772611	0.756408	0.929608
Word2Vec-BiLSTM (100)	0.747486	0.761992	0.754669	0.929658
Word2Vec-BiLSTM (50)	0.729093	0.759795	0.744128	0.927073
GloVe-BiLSTM (50)	0.738456	0.749542	0.743958	0.926576
FastText-BiLSTM (50)	0.714870	0.767484	0.740244	0.924886

Tablo XIII'te verilen genel sonuçlara göre BiLSTM-CRF mimarisinin statik kelime gömme yöntemleri ile çoğu durumda standart BiLSTM modeline göre daha yüksek performans sergilediği dikkat çekmektedir. Word2Vec-BiLSTM-CRF modelinin 100 vektör boyutunda 0.786401 F1 skoru ile CTI-NER görevinde yeterli anlamsal bilgi sağladığı ve CRF katmanının etiket geçişlerini başarılı şekilde modellediği görülmektedir.

Vektör boyutu açısından ele alındığında Word2Vec-BiLSTM-CRF modelinin, 50 vektör boyutunda 0.772942 F1 skoruna, 200 vektör boyutunda 0.778781 F1 skoruna, 300 vektör boyutunda 0.777881 F1 skoruna ve 100 vektör boyutunda 0.786401 F1 skoru ile en yüksek skora ulaşması her zaman daha büyük boyutlarının daha yüksek başarı getirmediğini kanıtlamaktadır.

Glove-BiLSTM-CRF modelinin 300 vektör boyutunda 0.779822 F1 skoru ile Word2Vec-BiLSTM-CRF modeline çok yakın performans göstermiştir. BiLSTM-CRF modelleri açısından 50 ve 300 vektör boyutlarında, BiLSTM modelleri açısından da 100, 200 vektör boyutların diğer modellere göre daha yüksek F1 skoru elde etmiştir. Bu durum, Glove kelime gömme yönteminin de küçümseyecek bir yöntem olmadığını göstermektedir.

FastText-BiLSTM-CRF modeli 200 vektör boyutunda 0.788722 geri çağırma değeri ile tablodaki en yüksek geri çağırma sonuçlarından birini üretmiştir. Bu durum FastText kelime gömme yönteminin, alt-kelime yapısı kullanması sayesinde teknik güvenlik terimlerini, hash değerlerini, exploit isimlerini ve nadir görülen CTI varlıklarını daha başarılı temsil edebilmesinden kaynaklanmaktadır [25]. Fakat kesinlik değerlerinin düşük olması, modelin daha fazla yanlış pozitif üretebildiğini göstermektedir.

C. Transformer Tabanlı Yöntemler

Transformer tabanlı yöntemlerin BiLSTM modeli ile sonuçları Tablo XIV'te, BiLSTM-CRF modeli ile sonuçları Tablo XV'te ve genel sonuçlar Tablo XVI'da sunulmuştur.

TABLO XIV. BILSTM SONUÇLARI

Model	Precision	Recall	F1	Accuracy
XLNet-BiLSTM	0.806911	0.833947	0.820206	0.943497
SecureBERT-BiLSTM	0.797694	0.838106	0.817401	0.940752
BERT-BiLSTM	0.803009	0.825783	0.814236	0.941845
RoBERTa-BiLSTM	0.796902	0.832047	0.814096	0.942380
ALBERT-BiLSTM	0.809169	0.797868	0.803479	0.940933

TABLO XV. BILSTM-CRF SONUÇLARI

Model	Precision	Recall	F1	Accuracy
SecureBERT-BiLSTM-CRF	0.825323	0.844714	0.834906	0.944038
XLNet-BiLSTM-CRF	0.820734	0.839470	0.829996	0.946190
RoBERTa-BiLSTM-CRF	0.821182	0.832047	0.826579	0.942829
BERT-BiLSTM-CRF	0.820522	0.821731	0.821126	0.946584
ALBERT-BiLSTM-CRF	0.804214	0.813671	0.808915	0.941382

TABLO XVI. GENEL SONUÇLAR

Model	Precision	Recall	F1	Accuracy
SecureBERT-BiLSTM-CRF	0.825323	0.844714	0.834906	0.944038
XLNet-BiLSTM-CRF	0.820734	0.839470	0.829996	0.946190
RoBERTa-BiLSTM-CRF	0.821182	0.832047	0.826579	0.942829
BERT-BiLSTM-CRF	0.820522	0.821731	0.821126	0.946584
XLNet-BiLSTM	0.806911	0.833947	0.820206	0.943497
SecureBERT-BiLSTM	0.797694	0.838106	0.817401	0.940752
BERT-BiLSTM	0.803009	0.825783	0.814236	0.941845
RoBERTa-BiLSTM	0.796902	0.832047	0.814096	0.942380
ALBERT-BiLSTM-CRF	0.804214	0.813671	0.808915	0.941382
ALBERT-BiLSTM	0.809169	0.797868	0.803479	0.940933

Tablo XVI incelendiğinde transformer tabanlı kelime gömme yöntemlerinin, CTI-NER görevinde klasik yöntemler ve statik kelime gömme yöntemlerine göre oldukça yüksek performans sergilediği görülmektedir. Bu durum, transformer tabanlı modellerin bağlamsal bilgiyi daha iyi öğrenebilme yeteneklerinden kaynaklanmaktadır [10].

SecureBERT-BiLSTM-CRF modeli, 0.834906 F1 skoru ile en yüksek performansı elde etmiştir. Elde edilen bu sonuç, siber güvenlik alanına yönelik olarak eğitilmiş SecureBERT modelinin CTI metinlerindeki terminolojiyi genel amaçlı transformer modellerine göre çok daha başarılı olarak öğrenebilmesinden ve bu sayede saldırı teknikleri, exploit isimleri ve tehdit aktörleri gibi siber güvenlik varlıklarını daha doğru sınıflandırmasından kaynaklanmaktadır [12].

Tablo XIV sonuçlarına göre, BiLSTM mimarisi açısından bakıldığında; XLNet-BiLSTM modeli 0.820206 F1 skoru elde ederken, SecureBERT-BiLSTM modeli 0.817401 F1 skorunda kalmıştır. Ancak CRF katmanının eklenmesiyle SecureBERT transformer modelinin performansındaki artışın XLNet modeline göre daha belirgin olduğu gözlemlenmiştir.

VI. SONUC VE TARTISMA

Bu çalışmada, yapılandırılmamış CTI metinlerinden NER görevi için 10 farklı kelime gömme yöntemi, DNRTI veriseti kullanılarak BiLSTM ve BiLSTM-CRF mimarileriyle test edilmiştir. Elde edilen test sonuçları, güncel doğal dil işleme literatüründeki siber güvenlik odaklı NER çalışmalarıyla karşılaştırıldığında güçlü bir tutarlılık ve ampirik geçerlilik sergilemektedir.

Tablo XVII'de 10 Farklı Kelime Gömme Yönteminin BiLSTM-CRF modeli ile test sonuçları, Tablo XVIII'de BiLSTM modelleri ile test sonuçları ve XIX'da en iyi performans gösteren ilk 5 modelin test sonuçları verilmiştir.

TABLO XVII. KELIME GÖMME YÖNTEMLERININ BILSTM-CRF SONUÇLARI

Model	Precision	Recall	F1	Accuracy
SecureBERT-BiLSTM-CRF	0.825323	0.844714	0.834906	0.944038
XLNet-BiLSTM-CRF	0.820734	0.839470	0.829996	0.946190
RoBERTa-BiLSTM-CRF	0.821182	0.832047	0.826579	0.942829
BERT-BiLSTM-CRF	0.820522	0.821731	0.821126	0.946584
ALBERT-BiLSTM-CRF	0.804214	0.813671	0.808915	0.941382
Word2Vec-BiLSTM-CRF (100)	0.793734	0.779202	0.786401	0.934778
GloVe-BiLSTM-CRF (300)	0.808812	0.752838	0.779822	0.931845
GloVe-BiLSTM-CRF (50)	0.784998	0.774075	0.779499	0.932044
Word2Vec-BiLSTM-CRF (200)	0.783543	0.774075	0.778781	0.931597
Word2Vec-BiLSTM-CRF (300)	0.788563	0.767484	0.777881	0.932591
GloVe-BiLSTM-CRF (200)	0.794039	0.760893	0.777113	0.929658
GloVe-BiLSTM-CRF (100)	0.790175	0.759795	0.774687	0.931845
FastText-BiLSTM-CRF (200)	0.759520	0.788722	0.773846	0.929111
Word2Vec-BiLSTM-CRF (50)	0.765720	0.780300	0.772942	0.929857
FastText-BiLSTM-CRF (300)	0.756947	0.787990	0.772156	0.926974
FastText-BiLSTM-CRF (50)	0.760073	0.766752	0.763398	0.927968
FastText-BiLSTM-CRF (100)	0.740663	0.777005	0.758399	0.925234

Model	Precision	Recall	F1	Accuracy
TFIDF-BiLSTM-CRF	0.767362	0.740388	0.753634	0.925482
BoW-BiLSTM-CRF	0.724415	0.692054	0.707865	0.913800

TABLO XVIII. KELIME GÖMME YÖNTEMLERININ BILSTM SONUÇLARI

Model	Precision	Recall	F1	Accuracy
XLNet-BiLSTM	0.806911	0.833947	0.820206	0.943497
SecureBERT-BiLSTM	0.797694	0.838106	0.817401	0.940752
BERT-BiLSTM	0.803009	0.825783	0.814236	0.941845
RoBERTa-BiLSTM	0.796902	0.832047	0.814096	0.942380
ALBERT-BiLSTM	0.809169	0.797868	0.803479	0.940933
Word2Vec-BiLSTM (300)	0.753965	0.783230	0.768319	0.932094
GloVe-BiLSTM (200)	0.781618	0.753570	0.767338	0.931298
FastText-BiLSTM (200)	0.770310	0.763823	0.767053	0.932342
GloVe-BiLSTM (300)	0.747624	0.777737	0.762383	0.929459
Word2Vec-BiLSTM (200)	0.749911	0.774075	0.761802	0.930603
FastText-BiLSTM (300)	0.745353	0.778103	0.761376	0.930006
GloVe-BiLSTM (100)	0.758278	0.754669	0.756469	0.929708
FastText-BiLSTM (100)	0.740871	0.772611	0.756408	0.929608
Word2Vec-BiLSTM (100)	0.747486	0.761992	0.754669	0.929658
Word2Vec-BiLSTM (50)	0.729093	0.759795	0.744128	0.927073
GloVe-BiLSTM (50)	0.738456	0.749542	0.743958	0.926576
FastText-BiLSTM (50)	0.714870	0.767484	0.740244	0.924886
TFIDF-BiLSTM	0.725712	0.727572	0.726641	0.921406
BoW-BiLSTM	0.719740	0.690223	0.704673	0.915242

TABLO XIX. EN İYI İLK 5 MODELIN SONUÇLARI

Model	Precision	Recall	F1	Accuracy
SecureBERT-BiLSTM-CRF	0.825323	0.844714	0.834906	0.944038
XLNet-BiLSTM-CRF	0.820734	0.839470	0.829996	0.946190
RoBERTa-BiLSTM-CRF	0.821182	0.832047	0.826579	0.942829
BERT-BiLSTM-CRF	0.820522	0.821731	0.821126	0.946584
XLNet-BiLSTM	0.806911	0.833947	0.820206	0.943497

Tablo XVII sonuçlarına göre; BoW ve TF-IDF gibi klasik yöntemler, BiLSTM-CRF modeli üzerinde sırasıyla 0.707865 ve 0.753634 F1 skorlarında kalmıştır. Word2Vec, GloVe ve FastText gibi statik kelime gömme yöntemleri ise bu başarıyı 0.76 – 0.78 bandına taşımıştır. Mevcut literatürde, siber güvenlik metinlerinin karmaşıklığını modellemek için sadece CNN, BiLSTM ve CRF mimarisi kullanan (bağlamsal kelime gömme yöntemleri kullanmayan) temel modellerin DNRTI verisetinde, 0.7501 gibi sınırda bir F1 skoru elde ettiği görülmektedir [2]. Çalışmada elde edilen statik yöntem sonuçları, geleneksel ve statik kelime gömme modellerinin siber saldırı zincirlerindeki bağlamsal ilişkileri çözmedeki yetersizliğine dair literatürdeki bu bulgularla tam bir uyum içindedir [2].

Tablo XVII'deki bağlama duyarlı transformatör modelleri incelendiğinde, performans belirgin bir sıçrama yaparak 0.82 XLNet F1 Skoru: 0.829996) seviyelerine çıkmıştır. Bu çalışmada; BERT-BiLSTM-CRF transformer tabanlı model 0.821126 F1 skoru ile Chen ve arkadaşları tarafından DNRTI veriseti üzerinde yapılan benzer çalışma ile elde edilen 0.8459 F1 skoruna yakın skor elde edilmiştir [4].

Literatürde Wang ve arkadaşlarının yaptığı çalışmada, XLNet-CRF mimarisi kullanarak CTI-Reports veri setinde 0.9743 gibi F1 skoru elde edilmesine karşın, DNRTI veri setinde 0.8239 F1 skorunda kalındığı tespit edilmiştir [5]. Yapılan çalışmada; DNRTI veriseti kullanılarak XLNet-BiLSTM-CRF modeli ile 0.829996 F1 skoru elde edilmiştir. Bu skor, Wang ve arkadaşlarının XLNET-CRF modeli ile elde ettiği skorun üzerine çıkmıştır.

Tablo XIX'daki sonuçlara göre; test edilen 10 kelime gömme yöntemi arasında en yüksek performans, 0.834906 F1 skoru ile doğrudan siber güvenlik verileriyle eğitilmiş olan SecureBERT-BiLSTM-CRF modelinde gözlemlenmiştir. Bu sonuç, literatürle yer alan Chen ve arkadaşlarının alana özgü olarak eğitilen SecBERT-BiLSTM-CRF modelinin DNRTI veriseti üzerinde ulaştığı 0.8377 F1 skoruna [4] ve Li ve arkadaşları tarafından geliştirilen SecureBERT_Plus kullanan CyberKG modelinin, aynı verisetinde elde ettiği 0.839 F1 skoruna [1] mükemmel bir şekilde yakınlık göstermektedir.

Tablo XVII ve Tablo XVIII tablolarındaki sonuçlar incelendiğinde; istisnasız tüm kelime gömme yöntemlerinde, modelin çıkışına CRF katmanı eklenmesinin F1 skorunu artırdığı görülmektedir. SecureBERT-BiLSTM modelinde 0.8174 olan F1 skoru, CRF eklenmesiyle 0.8349 değerine, XLNet-BiLSTM modelinde 0.820206 olan F1 skoru, CRF eklenmesiyle 0.829996 değerine çıkmıştır. Aynı şekilde klasik yöntemlerden BoW modelinde 0.7046'dan 0.7078'e, statik kelime gömme yöntemlerinden GloVe modelinde ise 0.7623'ten 0.7798'e yükseliş gözlemlenmiştir.

Çalışmadaki sonuçlar, literatürdeki güncel CTI-NER bulguları ile doğrulanmaktadır. Sonuç olarak; maksimum doğrulukla siber tehdit istihbaratı elde edebilmek için, alana özgü eğitilmiş bağlamsal dil modellerinin (SecureBERT) güçlü temsil yeteneklerinin, yapısal tutarlılığı sağlayan dizisel karar mekanizmalarıyla (BiLSTM-CRF) birleştirilmesi en optimal mimari çözümü sunmaktadır.

KAYNAKLAR

[1] B. Li, Q. Yang, C. Deng, and H. Pan, "CyberKG: Constructing a cybersecurity knowledge graph based on SecureBERT_Plus for CTI reports," Informatics, vol. 12, no. 3, p. 100, 2025.

[2] L. Chen, H. Deng, J. Zhang, B. Zheng, and R. Jiang, "Threat intelligence named entity recognition based on segment-level information extraction and similar semantic space construction," Symmetry, vol. 17, no. 5, p. 783, 2025.

[3] X. Feng, S. He, X. Wei, R. Liu, H. Yue, and X. Wang, "PROMPT-BART: A named entity recognition model applied to cyber threat intelligence," Applied Sciences, vol. 15, no. 18, p. 10276, 2025.

[4] S.-S. Chen, R.-H. Hwang, C.-Y. Sun, Y. R. Lin, and T.-W. Pai, "Enhancing cyber threat intelligence with named entity recognition using BERT-CRF," pp. 7532–7537, 2023.

[5] T. Wang, Y. Liu, C. Liang, B. Wang, and H. Liu, "XLNet-CRF: Efficient named entity recognition for cyber threat intelligence with permutation language modeling," Electronics, vol. 14, no. 15, p. 3034, 2025.

[6] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," arXiv preprint arXiv:1508.01991, 2015. [Online]. Available: https://arxiv.org/abs/1508.01991

[7] GeeksforGeeks, "Bag of Words vs TF-IDF," GeeksforGeeks, Jul. 23, 2025. [Online]. Available: https://www.geeksforgeeks.org/nlp/bag-of-words-vs-tf-idf/.

[8] C. Tulu, “Experimental comparison of pre-trained word embedding vectors of Word2Vec, GloVe, FastText for word level semantic text similarity measurement in Turkish,” Advances in Science and Technology – Research Journal, vol. 16, 2022.

[9] E. M. Dharma, F. L. Gaol, H. Leslie, H. S. Warnars, and B. Soewito, “The accuracy comparison among Word2Vec, GloVe, and FastText towards convolution neural network (CNN) text classification,” Journal of Theoretical and Applied Information Technology (JATIT), vol. 100, no. 2.

[10] S. F. N. Azizah, H. D. Cahyono, S. W. Sihwi and W. Widiarto, "Performance Analysis of Transformer Based Models (BERT, ALBERT, and RoBERTa) in Fake News Detection," 2023 6th International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2023, pp. 425-430

[11] J. Vitorino, E. Maia, and I. Praça, “Adversarial evasion attack efficiency against large language models,” arXiv preprint arXiv:2406.08050, 2024. [Online]. Available: https://arxiv.org/abs/2406.08050

[12] E. Aghaei, X. Niu, W. Shadid, and E. Al-Shaer, “SecureBERT: A domain-specific language model for cybersecurity,” arXiv preprint arXiv:2204.02685, 2022. [Online]. Available: https://arxiv.org/abs/2204.02685

[13] D. Kurniasari, F. A. Pradana, and W. Wamiliana, “Deep learning-driven comparative study of word embedding techniques: Word2Vec, GloVe, and FastText in health condition reviews,” ISRG Journal of Multidisciplinary Studies (ISRGJMS), 2025. [Online]. Available: http://repository.lppm.unila.ac.id/54439/

[14] H. S. Shin, H. Y. Kwon, and S. J. Ryu, “A new text classification model based on contrastive word embedding for detecting cybersecurity intelligence in Twitter,” Electronics, vol. 9, no. 9, p. 1527, 2020.

[15] M. A. Haq, M. A. R. Khan, and M. Alshehri, “Insider threat detection based on NLP word embedding and machine learning,” Intelligent Automation & Soft Computing, 2022.

[16] P. Evangelatos, C. Iliou, T. Mavropoulos, K. Apostolou, T. Tsikrika, S. Vrochidis, and I. Kompatsiaris, “Named entity recognition in cyber threat intelligence using transformer-based models,” in Proc. IEEE Int. Conf. on Cyber Security and Resilience (CSR), 2021.

[17] E. Aghaei, S. Jain, P. Arun, and A. Sambamoorthy, “SecureBERT 2.0: Advanced language model for cybersecurity intelligence,” in 2025 Annual Computer Security Applications Conference Workshops (ACSAC Workshops). IEEE, 2025, pp. 382–388.

[18] L. Ye, Y. Wu, H. Zhang, and M. Ge, “A cybersecurity NER method based on hard and easy labeled training data discrimination,” Sensors, vol. 25, no. 24, p. 7627, 2025.

[19] X. Wang, X. Liu, S. Ao, N. Li, Z. Jiang, and Z. Xu, “DNRTI: A large-scale dataset for named entity recognition in threat intelligence,” in Proc. IEEE Int. Conf. on Trust, Security and Privacy in Computing and Communications (TrustCom), 2020.

[20] Y. Mudugandla, “Word embeddings in NLP: From bag of words to transformers – Part 1,” Towards AI, [Online]. Available: https://towardsai.net/p/machine-learning/word-embeddings-in-nlp-from-bag-of-words-to-transformers-part-1

[21] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in Advances in Neural Information Processing Systems, vol. 26, 2013.

[22] SAP Community, “GloVe and FastText: Two popular word vector models in NLP,” SAP Community, [Online]. Available: https://community.sap.com/t5/artificial-intelligence-blogs-posts/glove-and-fasttext-two-popular-word-vector-models-in-nlp/ba-p/13473570

[23] R. Řehůřek, “Word2Vec tutorial,” Gensim Documentation, [Online]. Available: https://radimrehurek.com/gensim/auto_examples/ tutorials/ run_word2vec.html

[24] J. Pennington, R. Socher, and C. Manning, “GloVe: Global vectors for word representation,” in Proc. Conf. on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1532–1543.

[25] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” arXiv preprint arXiv:1607.04606, 2017. [Online]. Available: https://arxiv.org/abs/1607.04606

[26] R. Řehůřek, “FastText tutorial,” Gensim Documentation, [Online]. Available: https://radimrehurek.com/gensim/auto_examples/tutorials/ run_fasttext.html

[27] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint arXiv:1810.04805, 2019. [Online]. Available: https://arxiv.org/abs/1810.04805

[28] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, “RoBERTa: A robustly optimized BERT pretraining approach,” arXiv preprint arXiv:1907.11692, 2019. [Online]. Available: https://arxiv.org/abs/1907.11692

[29] T. B. Faruk, “Automated CVE analysis: Harnessing machine learning in designing question-answering models for cybersecurity information extraction,” arXiv preprint arXiv:2412.16484, 2024. [Online]. Available: https://arxiv.org/abs/2412.16484

[30] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, “ALBERT: A lite BERT for self-supervised learning of language representations,” in Proc. Int. Conf. on Learning Representations (ICLR), 2020. [Online]. Available: https://arxiv.org/abs/1909.11942

[31] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, “XLNet: Generalized autoregressive pretraining for language understanding,” in Advances in Neural Information Processing Systems (NeurIPS), 2019. [Online]. Available: https://arxiv.org/abs /1906.08237

[32] Y. Ech-Chammakhy, A. Motii, A. Rabii, O. Azrara, and J. Chbili, “CyberNER: A harmonized STIX corpus for cybersecurity named entity recognition,” arXiv preprint arXiv:2510.26499, 2025. [Online]. Available: https://arxiv.org/abs/2510.26499

[33] S. Khatoon, J. Mir, and A. Mahmood, “Named entity recognition with feature-enhanced BiLSTM and CRF for fine-grained aspect identification in large-scale textual reviews,” Machine Learning and Knowledge Extraction, vol. 8, no. 4, p. 88, 2026.

[34] N. Manchev, “Named entity recognition (NER): Challenges and model,” Domino Data Lab Blog, [Online]. Available: https://domino.ai/ blog/named-entity-recognition-ner-challenges-and-model

[35] J. Lafferty, A. McCallum, and F. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” in Proc. 18th Int. Conf. on Machine Learning (ICML), 2001, pp. 282–289.

Bulut Tabanlı Akıllı Şehir Altyapılarında Yapay Zekâ Destekli Ortadaki Adam Saldırılarının Tespiti ve Önlenmesi

Halis AYDOĞAN[1], Cevat ÖZARPA[2]

1.Ankara Medipol Üniversitesi Fen Bilimleri Enstitüsü Siber Güvenlik ABD

(ORCID ID: 0009-0003-2098-4721)
halis.aydogan@ankaramedipol.edu.tr

2. Ankara Medipol Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi Biyomedikal Mühendisliği Bölümü
(ORCID ID: 0000-0002-1195-2344)
cevat.ozarpa@ankaramedipol.edu.tr

Özet —Ortadaki Adam (MitM) saldırıları, büyük ölçekli veri alışverişinin heterojen ve kaynak kısıtlı cihazlar arasında gerçekleştiği bulut tabanlı akıllı şehir altyapılarında iletişim bütünlüğü için kritik bir tehdit oluşturmaktadır. Nesnelerin İnterneti (IoT) sistemlerinin ve bulut merkezli mimarilerin hızla yayılması, saldırı yüzeyini önemli ölçüde genişleterek, DNS sahtekarlığı, SSL soyma, ARP zehirlenmesi ve sahte erişim noktası istismarı gibi daha karmaşık ve koordineli MitM saldırılarına olanak sağlamıştır.

Bu çalışma, bulut entegre akıllı şehir ortamlarında MitM saldırılarını belirlemek için yapay zekâ tabanlı bir tespit ve azaltma çerçevesi önermektedir. Önerilen yaklaşım, ağ trafiği modellerini izlemek ve müdahale veya manipülasyon girişimlerini gösteren sapmaları belirlemek için makine öğrenimine dayalı anomali tespitinden yararlanmaktadır.

Dinamik ve büyük ölçekli altyapılarda yanlış pozitifleri en aza indirirken tespit doğruluğunu artırmak için risk odaklı bir model geliştirilmiştir.

Çerçeve, simüle edilmiş akıllı şehir ağ senaryoları aracılığıyla değerlendirilmiş ve geleneksel kural tabanlı güvenlik mekanizmalarına kıyasla iyileştirilmiş tespit performansı ve dayanıklılık göstermiştir. Sonuçlar, yapay zekâ tabanlı yaklaşımların gerçek zamanlı tehdit tespitini önemli ölçüde geliştirebileceğini, kritik altyapılarda veri bütünlüğünü, gizliliğini ve hizmet sürekliliğini sağlayabileceğini göstermektedir.

Bulgular, mevcut bulut güvenlik mimarilerindeki temel sınırlamaları ele alarak ve gelişen MitM tehditlerine karşı ölçeklenebilir bir savunma modeli sağlayarak, yeni nesil akıllı şehirler için akıllı ve uyarlanabilir siber güvenlik çözümlerinin geliştirilmesine katkıda bulunmaktadır.

Anahtar Kelimeler—Ortadaki Adam (MitM), Bulut Güvenliği, Akıllı Şehirler, Yapay Zekâ, Anomali Tespiti, IoT Güvenliği, Ağ Güvenliği

Abstract —Man-in-the-Middle (MitM) attacks pose a critical threat to communication integrity in cloud-based smart city infrastructures where large-scale data exchange occurs between heterogeneous and resource-constrained devices. The rapid proliferation of Internet of Things (IoT) systems and cloud-centric architectures has significantly expanded the attack surface, enabling more sophisticated and coordinated MitM attacks such as DNS spoofing, SSL snatching, ARP poisoning, and spoofed access point exploitation.

This study proposes an AI-based detection and mitigation framework for identifying MitM attacks in cloud-integrated smart city environments. The proposed approach leverages machine learning-based anomaly detection to monitor network traffic patterns and identify deviations indicating attempted interference or manipulation. A risk-driven model has been developed to increase detection accuracy while

minimizing false positives in dynamic and large-scale infrastructures.

The framework has been evaluated through simulated smart city network scenarios and has demonstrated improved detection performance and resilience compared to traditional rule-based security mechanisms. The results show that AI-based approaches can significantly improve real-time threat detection, ensuring data integrity, confidentiality, and service continuity in critical infrastructure.

The findings contribute to the development of intelligent and adaptable cybersecurity solutions for next-generation smart cities by addressing key limitations in current cloud security architectures and providing a scalable defense model against evolving Man-in-the-Middle (MitM) threats.

Keywords—Man-in-the-Middle (MitM), Cloud Security, Smart Cities, Artificial Intelligence, Anomaly Detection, IoT Security, Network Security

I. GİRİŞ

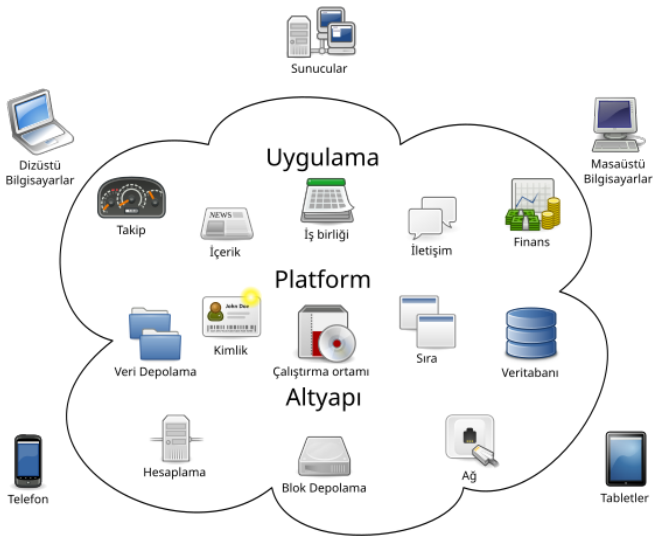
Bulut bilişim, internet üzerinden erişim imkânı sunan teknolojik bir gelişmedir. Bulut bilişim, internetin olduğu her yerde erişim olanağı sağlar. Bulut bilişim güvenlik sorunu ise kullanıcıları çok fazla olumsuz etkileyecek düzeyde bir sorundur.[1] Man in the Middle güvenlik saldırılarının başında gelmektedir. Man-in-the-Middle (MitM) saldırıları ve bulut bilişimde güvenlik zafiyetleri, siber güvenlik dünyasında önemli kritik bir yere sahiptir. Teknolojinin hızla geliştiği bu zamanlarda, veri transferi ve depolama yöntemlerinde büyük değişimler gözle görülür halde gelişmektedir. Ancak bu gelişmeler, yeni güvenlik tehditlerini ve güvenlik zafiyetlerini de beraberinde getirmektedir. Bu makalede, MitM saldırılarının tanımı, türleri, önleme yöntemleri ve bulut bilişimdeki güvenlik açıkları ele alınacaktır.

II. BULUT BİLİŞİMDE GÜVENLİK AÇIKLARI

Bulut bilişimdeki veri güvenliği, bulut ortamlarında depolanan, işlenen ve iletilen hassas verilerin korunmasını ifade eder. Yetkisiz erişime, veri ihlallerine ve diğer tehditlere karşı koruma sağlamak için çeşitli güvenlik önlemlerinin uygulanmasını içerir. Bulut bilişimdeki veri güvenliğinin temel yönlerinden biri, güvenlik açıklarının ve yanlış yapılandırmaların belirlenmesi ve azaltılmasıdır.

A. Bulut Bilişimin Tanımı ve Avantajları

Bulut bilişim, depolama, sunucu, veri tabanı, ağ, yazılım gibi bilgi işlem hizmetlerinin internet üzerinden kesinti yaşanmadan erişimin sağlanmasıdır. İngilizce karşılığı "Cloud Computing" olan bulut teknolojisi sanal sunucularda depolanan verilerimize tüm cihazlardan ve istenilen her yerden sorunsuz bir şekilde kolaylıkla ulaşmamızı mümkün kılar.[2] Günümüzde gelişmeye ve geliştirmeye açık olan bulut teknolojileri birden fazla şirket hizmet vermektedir. Kullanıcıların bulut bilişimi kullanırken yapacağı şirket sözleşmelerini dikkate alıp kullanım şartlarını iyi okuyup belirlemeleri önemlidir. Bulut teknolojisi, esneklik, maliyet tasarrufu ve yüksek erişilebilirlik gibi avantajlar sunar. Ancak, bu faydaların yanında güvenlik riskleri de bulunmaktadır.



Şekil 1. Bulut bilişim

III. BULUT BİLİŞİMDEKİ TEMEL GÜVENLİK RİSKLERİ

Bulut bilişimde temel güvenlik açıkları ve riskleriyle, işletme açısından hassas bilgiler açığa çıkabilir ve işletme operasyonlarını etkileyebilir.

A. Veri Gizliliği ve Çalıntı Veriler

Veri gizliliği kurumsal veya kişisel verilerin izinsiz erişimlere, hırsızlığa, açıklamalara karşı korunmasını ifade etmektedir. Bulut bilişimde veri gizliliği, hassas bilgiler (örneğin, müşteri bilgileri, finansal veriler, şirket sırları gibi) bulut hizmet sağlayıcıları tarafından saklandığı ve işlendiği durumlar için büyük önem taşır. Bulut ortamında depolanan veriler, çeşitli saldırılara karşı hassastır. İç ağ; ev, ofis, laboratuvar, bina vb. küçük ölçekli bir yerel ağdır. İç ağlarda özellikle kablosuz erişim cihazlarından veri dinlenmesi kablolu erişimde kullanılan anahtar cihazlara göre daha kolaydır[3]. Yetersiz şifreleme veya zayıf kimlik doğrulama sistemleri bu riski arttırabilir. Veri gizliliği ile ilgili diğer tehditler ise; yetkisiz erişim, bulut sağlayıcı güvenliği, veri işleme ve paylaşımı, çok kiracılı ortamlar olarak örneklendirilebilir.[4] Veri gizliliğini sağlama yöntemleri ise, güçlü şifreleme kullanımı, erişim kontrolleri, hassas veri yönetimi, şeffaflık ve sözleşme güvencesi gibi maddelerle gösterilebilir.

B. Hizmet Kesintileri

Bulut hizmet sağlayıcılarındaki arızalar veya siber saldırılar, hizmet kesintilerine yol açabilir. Hizmet kesintisi, bulut hizmet sağlayıcısının sistemlerinde yaşanan bir arıza ya da farklı sebepler nedeniyle kullanıcıların verilerini veya uygulamalara erişiminin engellenmesi durumudur. Kesinti süreleri birkaç dakikadan saatlere kadar sürebilir ve sistem tamamen çevirim dışı kalabilir. Bu durumda da bazı işlevler kısıtlanabilir ve tamamen erişim olmayabilir.

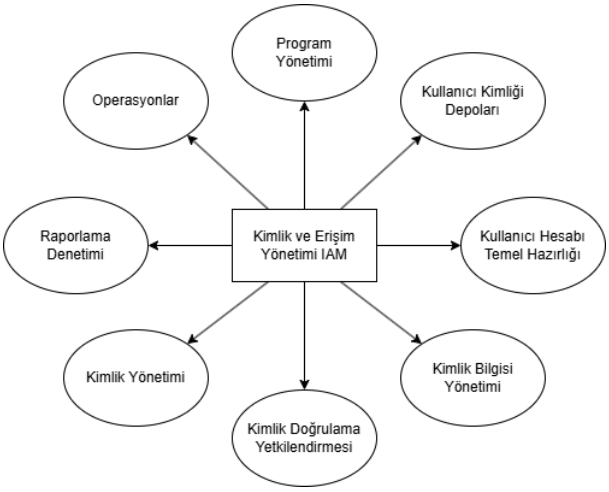
Hizmet kesintilerinin sebepleri; teknik arızalar, ağ sorunları, doğal afetler, siber saldırılar, enerji kesintileri, hizmet sağlayıcı hatalar ve planlı bakımlar olarak nitelendirilebilir. Bu hizmet kesintilerinin sonucunda, iş süreçlerini aksaması, finansal kayıplar, müşteri memnuniyetsizliği, işletme güvensizliği, veri kaybına ve hukuki ve yasal sorunlara neden olabilir.

Hizmet kesintilerini azaltma ve yönetme yöntemleri olarak, yedekleme ve felaket kurtarma planları yapılmalıdır, çok veri merkezi kullanılmalıdır, izleme ve erken uyarı sistemleri kullanılmalıdır, SLA taahhütleri gözden geçirilip kesinti süresinin anlaşmada verildiği süreyi geçtiği anda tazminat gibi koruma maddeleri içermelidir. Planlı bakım süreleri kullanıcılara önceden paylaşılmalıdır. Yük dengeleme (Load Balancing) ani trafik artışında sistemi koruması adına kullanılmalıdır. Bulut sağlayıcılarının, güvenlik önlemleri kesinti geçmişi dikkate alınmalıdır. Alternatif bulut sağlayıcılarla kritik hizmetler için birden fazla kullanılarak yedeklilik sağlanmalıdır.

IV. KİMLİK ve ERİŞİM YÖNETİMİ (IAM) ZAFİYETLERİ

IAM zafiyetleri, kimlik doğrulama veya erişim kontrol mekanizmalarındaki açıklar nedeniyle yetkisiz erişimlere veya güvenlik ihlallerine yol açabilmektedir. Örnek olarak, zayıf veya tahmin edilebilir parolalar, gereğinden fazla yetkilendirme, yetkilendirme kontrollerinin eksikliği, kimlik doğrulama saldırıları, erişim loglarının ve izleme sistemlerinin eksikliği, eski veya devre dışı bırakılmamış kullanıcı hesapları, oturum yönetimi zafiyetleri, kötü yapılandırılmış IAM politikaları, sosyal mühendislik saldırıları, çok faktörlü kimlik doğrulama eksiklikleri olarak belirtilebilir.[5]

IAM sistemleri, güvenliğin ilk savunma mekanizmasıdır. Bu sistemlerdeki zafiyetler büyük riskler oluşturur. Güçlü parola politikaları, çok faktörlü kimlik doğrulama ve düzenli erişim denetlemeleri gibi önlemler bu zafiyetlerin etkisini büyük ölçüde azaltabilir.



Şekil 2. Kimlik ve erişim yönetimi (IAM)

V. VERİ KAYBI

Veri kaybı, dijital veya fiziksel verilerin, kullanıcı veya sistemin tekrar erişemeyeceği şekilde kaybolması, silinmesi veya bozulmasıdır. Bu kayıp, bireysel kullanıcılardan büyük şirketlere kadar herkes için ciddi sonuçlar yaratabilir. Veri kaybı hem teknik hatalar hem de kötü niyetli faaliyetler sonucu gerçekleşebilir. Veri kaybının nedenleri olarak, siber saldırılar, donanım arızaları, insan hataları, yazılım hataları, doğal afetler, veri taşımada kaybolma, yetersiz yedekleme gibi örnekler verebiliriz. Veri kaybı hem bireysel hem de kurumsal düzeyde ciddi sonuçlar doğurabilen bir durumdur. Güçlü yedekleme stratejileri, siber güvenlik önlemleri ve kullanıcı farkındalığı, veri kaybı riskini büyük boyutta azaltabilir. İyi bir şekilde hazırlanmış veri yönetimi planı, kayıpları önlemek ve kurtarma süreçlerini kolaylaştırır ve hayati önem taşır.[6]

VI. BULUT BİLİŞİMDE GÜVENLİK ÖNLEMLERİ

Bulut bilişimde güvenlik, verilerin ve hizmetlerin bulut ortamında korunmasını sağlamak için alınan önlemleri kapsar. Siber saldırılarının sayısı ve siber saldırı mağdurlarının ekonomik açıdan kaybı hakkında kesin bilgiler elde etmek neredeyse imkânsızdır, çünkü suçlular nadiren tespit edilmekte ya da yetkili kurumlara bildirilmemektedir[7].

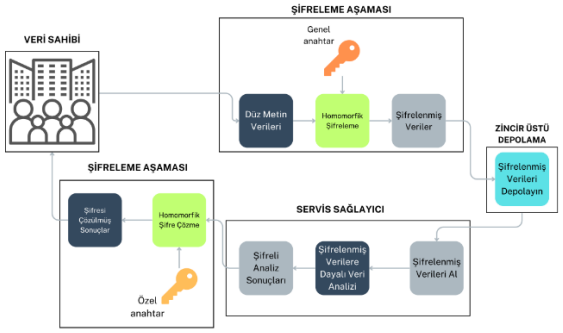
A. Şifreleme

Verilerin hem aktarım hem de depolama sırasında şifrelenmesi, gizliliği sağlar. Şifreleme, verilerin, yalnızca yetkili tarafların erişebilmesi için bir algoritma kullanılarak okunamaz hale getirilmesidir. Bulut bilişimde şifreleme, verilerin hem depolama sırasında (at rest) hem de transfer sırasında (in transit) korunmasını sağlar. Bu yöntem, verilerin güvenliğini artırır ve siber saldırılara karşı bir savunma hattı oluşturur. Şifreleme türleri olarak, veri aktarımında şifreleme, depolama sırasında şifreleme, istemci tarafında şifreleme, homomorfik şifreleme olarak gösterilebilir.

Şifrelemenin avantajları olarak, veri gizliliği verileri yetkisiz kişilerden korur. Yasal uyum ile veri koruma yasalarına uyumu desteklemektedir. Güvenlik ihlallerini önleyerek saldırganların veri çalması durumunda bilgilere erişimi engeller. Müşterilerin verilerin güvenliğini sağladığı için güven oluşturur.[8]

Şifreleme türleri ve anahtar yönetimi stratejileri, farklı kullanım durumlarına göre dikkatle seçilmelidir. Güvenli bir bulut ortamı oluşturmak için şifreleme politikalarının doğru bir şekilde uygulanması gereklidir.

Ortadaki adam saldırısını engellemenin bir diğer yolu dijital imzaları kullanabilen bir açık anahtar kripto sistemi kullanmaktır. Kurulum için her iki tarafta karşı tarafın açık anahtarını bilmelidir. Paylaşılan gizlilik oluşturulduktan sonra, taraflar kendi dijital imzalarını karşı tarafa göndermelidir. Ortadaki adam bu imzaları taklit etmeye çalışacak, fakat imzaların sahtesini yapamayacağı için başarısız olacaktır[9].



Şekil 3. Homografik Şifreleme

B. Kimlik ve Erişim Yönetimi

Çok faktörlü kimlik doğrulama ve rol tabanlı erişim kontrolleri, yetkisiz erişimleri engelleyebilir. Kimlik ve Erişim Yönetimi (IAM), kullanıcıların ve sistemlerin bir organizasyonun kaynaklarına erişimini güvenli bir şekilde kontrol etmek için kullanılan bir sistemdir. Amaç, yalnızca yetkili kişilerin doğru kaynaklara erişebilmesini sağlamak ve yetkisiz erişimi engellemektir.

IAM, kimlik doğrulama ve yetkilendirme olmak üzere ikiye ayrılmaktadır. Avantajları olarak, merkezileştirilmiş erişim yönetimi, güçlü kimlik doğrulama, yetki yönetimi, izleme ve analiz örnek verilebilir.

Kimlik ve Erişim Yönetimi, siber güvenlikte kritik bir rol oynar. Modern IAM çözümleri, güçlü kimlik doğrulama yöntemleri, merkezi yönetim ve otomatik izleme özellikleriyle organizasyonların güvenlik seviyesini artırır. Özellikle bulut bilişim ortamlarında, IAM' in doğru uygulanması hem güvenlik hem de operasyonel verimlilik açısından büyük önem taşır.

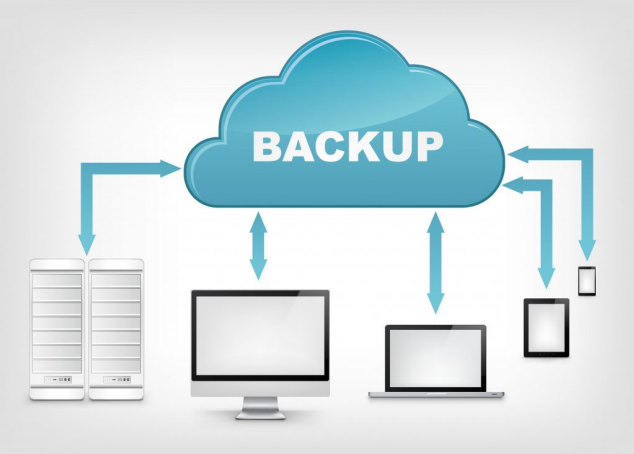
C. Güvenlik Denetimleri

Bulut sağlayıcılarının düzenli olarak güvenlik denetimlerinden geçirilmesi, zafiyetleri azaltabilir. Sistem ve süreçlerin güvenlik politikalarına uygun olup olmadığını kontrol etmek ve olası güvenlik açıklarını tespit etmek amacıyla yapılan değerlendirmelerdir.[10] Güvenlik denetim türleri, uyumluluk denetimi, teknik güvenlik denetimi, operasyonel güvenlik denetimi, iç ve dış güvenlik denetimleri olarak örnek verilebilir.

Güvenlik denetiminin, risk yönetimi, uyumluluk, operasyonel verimlilik, güven artışı gibi sağladığı avantajlar bulunmaktadır. Özellikle bulut bilişimde, düzenli denetimler sayesinde potansiyel zafiyetler tespit edilip giderilebilir ve güvenlik politikalarına uygunluk sağlanabilir.

D. Yedekleme

Yedekleme, verilerin bir kopyasının farklı bir konuma veya cihaza kaydedilmesi işlemidir. Bu, veri kaybı durumunda (örneğin, donanım arızası, siber saldırı, insan hatası veya doğal afetler) verilerin geri yüklenmesini sağlar. Veri kaybı riskine karşı düzenli ve güvenli yedekleme stratejileri uygulanmalıdır.



Şekil 4. Yedekleme

Tablo 1. *Kritik Öneme Sahip Durumlar*

- Fidye yazılım saldırıları sonrası veri kurtarma.
- Donanım arızalarından sonra sistemlerin çalıştırılması.
- Doğal afetler veya yangın gibi felaketlerde veri kaybını önleme.

Yedekleme, iş sürekliliği ve afet kurtarma planlarının temel bir bileşenidir. Yukarıdaki Tablo1'de ki gibi durumlar için kritik öneme sahiptir.[8]

E. Eğitim ve Farkındalık

Sosyal mühendislik saldırılarına karşı savunmalarda, phishing, baiting ve diğer sosyal mühendislik saldırıları, kullanıcıların bilinçsizliği nedeniyle başarılı olur. Eğitim, bu tür saldırılara karşı koruma sağlar ve saldırıları önleme konusunda yardımcı olur. Güvenlik politikalarının uygulanması ise kullanıcıların güvenlik protokollerine uyumunu arttırır ve zafiyetlerin azalmasına olanak sağlar. İnsan hatasının azalmasıyla yanlış yapılandırma, zayıf parolalar veya yanlışlıkla hassas verilerin sızdırılması gibi kullanıcı kaynaklı hatalar önlenmiş olunur.

Eğitim için yöntem ve araçlar olarak, farkındalık kampanyaları düzenlenmelidir. Bunlar, düzenli bilgilendirme e-postaları, görsellere yer verilen posterler ve materyaller kullanılmalıdır. Online ve fiziksel eğitimler verilmelidir. Sahte saldırı simülasyonları kullanıcıların durumu göz ardı etmesinin önüne geçer. Oyunlaştırarak çalışanları motive edip doğru güvenlik adımlarını uygulayanlar ödüllendirilebilir.

Eğitim ve farkındalık programlarının başarısı için, üst düzey yöneticilerin katılımı ve desteği programların ciddiye alındığını gösterir, Sürekli güvenlik tehditleri ile ilgili eğitimler düzenlenmelidir. Eğitimlerin etkisi düzenli olarak ölçülmeli sürekli değişip gelişen saldırılar için zayıf noktalar tespit edilmelidir.[11] Teknik bilgi seviyesi kullanıcılar arasında farklılık göstereceği için her seviyeden çalışanlar için uygun içerikler hazırlanmalıdır.

Eğitim konularına örnek ise, siber güvenlik farkındalığı, veri koruma analizi ve gizliliği, bulut güvenlik politikaları, iş sürekliliği ve felaket kurtarma olabilir.

Eğitim ve farkındalık, güvenliğin zayıf halkası olarak görülen insan faktörünü güçlü bir savunma hattına dönüştürmek ve bilinçli kullanıcılar yetiştirmek, gelecek saldırılara karşı savunmalar için önemlidir.

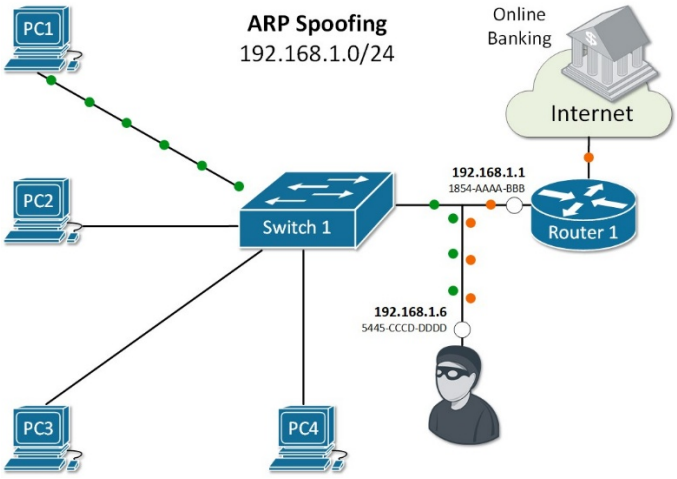
VII. MitM SALDIRILARININ TANIMI

MitM saldırıları, iki taraf arasındaki iletişime izinsiz olarak araya girmesi müdahale edilmesi ve bilgi çalması durumudur. Saldırgan, hedeflerin fark etmeyeceği bir şekilde iletişimi dinleyebilir, görebilir, değiştirebilir, manipüle edebilir veya kaydedebilir. MitM saldırıları, başarıyla uygulanması halinde saldırgana büyük avantajlar sağlar. Bu saldırı türü oldukça yaygın olmakla birlikte en az önlem alınan saldırılar arasındadır[12]. Bu tür saldırılar genellikle şu amaca hizmet eder:

- Gizli bilgilere erişim.
- Veri manipülasyonu.
- Kimlik hırsızlığı.

VIII. Man-in-the-Middle (MitM) SALDIRILARI

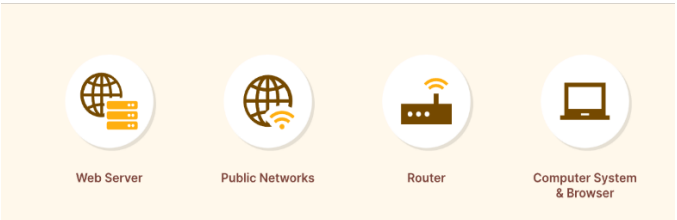
Ortadaki adam saldırıları bulut bilişime yapılan en önemli saldırılar arasında belirtilmektedir. Bulut bilişimde, sisteminizdeki açıklardan yararlanılarak zafiyetler kullanılır.



Şekil 5. Man in the Middle saldırıları

A- MitM Saldırılarının Türleri

MitM saldırıları, bir saldırganın iki taraf arasındaki iletişime gizlice dahil olarak veri elde etmesini, manipüle etmesini veya yönlendirmesini hedefler. Saldırı olduğunda iletişimde olan kurban düğümlerden olan mesaj gönderen kaynak düğümün durumu "Sender", mesajı alan hedefin durumu "Dest" ve araya girip paketleri üzerinden geçiren saldırgan düğümün durumu "MitM" olarak belirtilmiştir[13].

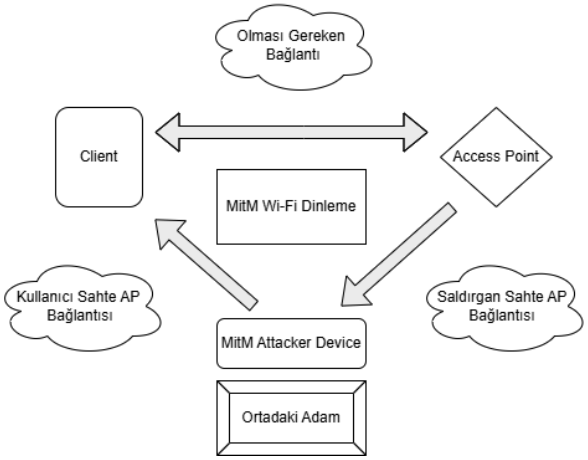


Şekil 6. Saldırı yerleri

1) Wi-Fi Dinleme (Evil Twin)

Saldırgan, meşru bir Wi-Fi ağını taklit ederek kullanıcıları sahte bir ağa bağlanması için kandırabilir. Bu sayede, kullanıcının girdiği tüm veriler ele geçirilebilir. Bu saldırı gerçekleştirmesi zor bir saldırı olmasına rağmen günümüzde fazlasıyla etkili olmaktadır.

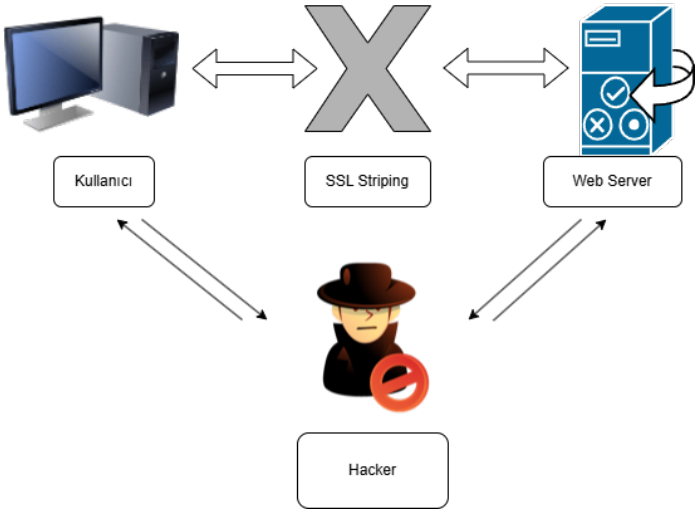
Wi-Fi saldırılarına karşı şifrelemeleri güvenli ve uzun olarak karakter kullanılarak uygulanmalıdır.



Şekil 7. MitM Wi-Fi Dineleme

2) SSL Striping

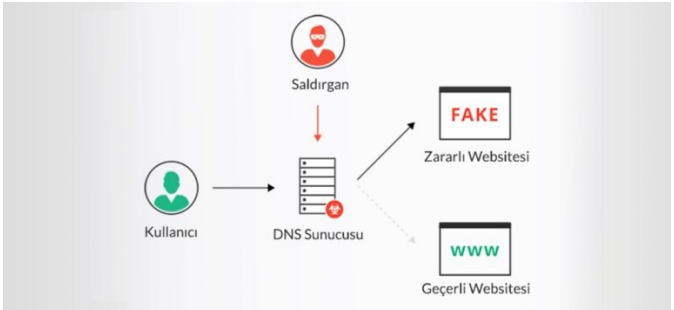
SSL Striping saldırılarında, saldıran kişi şifreli bir bağlantıyı şifresiz bir hale dönüştürür ve kullanıcının verilerini ele geçirir. SSL bilgiyi ele geçirme saldırısında, saldırgan istemci ile sunucu arasındaki tüm trafiği kestikten sonra ve istemcinin herhangi bir SSL içeriğini ele geçirip sunucuya iletir.



Şekil 8. SSL Striping

3) DNS Spoofing

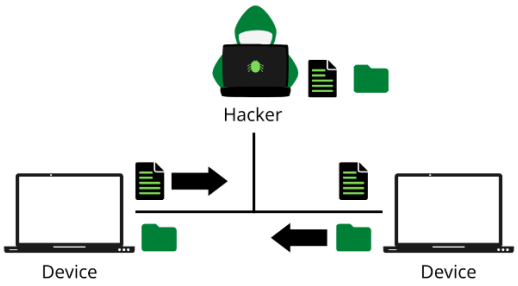
MitM saldırıları, DNS sunucularını da hedefleyebilir. DNS spoofing ile saldırgan, kullanıcıları sahte web sitelerine yönlendirmek için DNS yanıtlarını manipüle eder. DNS araması, web tarayıcılarının kendi adlarını IP adreslerine dönüştürür ve web sitelerine ulaşmasını sağlar. DNS sahtekarlığı ve DNS ele geçirme gibi MitM saldırılarında, genellikle kötü amaçlı yazılımı öne sürüp önemli bilgiler toplayabilir[14]. MitM saldırıları genellikle bankacılık ve e-ticaret sitelerini hedef alır ve bu durum da izinsiz giriş yapmaya çalışan kullanıcıların kimlik bilgilerini ve değerli bilgileri ele geçirmesine neden olabilir.[15]



Şekil 7. DNS saldırıları

4) ARP (Adres Çözümleme Protokolü) Spoofing

ARP saldırıları, yerel ağda veri paketlerinin saldırgana yönlendirilmesi için ARP tablosunun manipüle edilmesini içerir. Elektronik cihazlarımızda iki tür adres bulunmaktadır. Bunlar IP adres ve MAC adresleridir. MAC tabanlı doğrulama 802.11 standardında tanımlanmamıştır. Bununla birlikte birçok tedarikçi MAC tabanlı doğrulamayı uyarlamıştır[16]. Bir yerel alan ağından (LAN) gelen trafikte MAC adresinden bilgileri öğrenmeye çalışır, bilgi alamadığı cihazlarda ise IP adresine giderek oradan bilgi almak ister. Bir saldırgan ise sahte MAC adresi ile araya girdikten sonra ağda ki trafiğe sahte MAC adresi ekleyebilir trafiği izleyebilir ve kötü amaçlı kullanabilir.[17] MAC Adres Taşması en çok yapılan 2. Katman saldırısıdır[18].



Şekil 8. Adres çözümleme Protokolü

9. BULUT SİSTEMLERDE AKILLI ŞEHİR YAPISI VE MITM SALDIRILARI

Akıllı şehirler; trafik yönetiminden enerji dağıtımına, atık yönetiminden acil durum müdahale sistemlerine kadar geniş bir yelpazede sunmaktadır. IoT sensörleri, kameralar ve aktüatörler kullanmaktadır.[19] Bulut bilişim de ise bu cihazların ürettiği büyük boyutlu verilerin depolanmasını, anlamlandırmasını ve işlenmesini sağlamaktadır. Bulut, akıllı şehirlere maaliyet ve ölçeklenebilirlik faydası sunarken beraberinde de siber saldırılara karşı koruma oranını da yükseltmektedir.

Ortadaki Adam (MITM) saldırıları, bir saldırganın iki taraf (örneğin bir akıllı trafik ışığı ile bulut sunucusu) arasındaki iletişime gizlice girerek, tarafların birbirleriyle doğrudan haberleştiklerini zannederken, saldırganın tüm veri akışını okuyabildiği, değiştirebildiği veya yönlendirebildiği bir siber tehdit türüdür.[15] Akıllı şehir kapsamında yapılan bu saldırılar, bulut sistemlerde ve bağlantılarında yıkıcı sonuçlar ortaya koyabilir. Saldırgan, veri gizliliği ihlali, veri bütünlüğünün bozulması ve kontrolü ele geçirme gibi fiziksel yapıyı doğrudan manipüle edebilir. Yapılan literatür taramaları, akıllı şehirlerdeki yüksek hızlı ağlar ve bulut sistemlerinin MITM, DDoS ve Sybil gibi çok çeşitli pasif ve aktif saldırılara maruz kaldığını göstermektedir.[20] MITM saldırıları, zayıf şifreli ve halka açık Wi-Fi üzerinden bulut sistemlere bağlanan cihazlar, teknik olarak yapılan saldırıların tespit edilmesi oldukça zordur. Ağ Segmentasyonu ve izleme; akıllı şehir ağları, bulut bağlantıları mantıksal olarak ayrıştırılmalı ve anormal trafik modelleri (ARP zehirlenmesi, DNS spoofing gibi MITM belirtileri) için sürekli izlenmelidir.[21] Güçlü şifreleme ve kimlik doğrulama, VPN kullanılması, yapay zekâ destekli tehdit tespiti yapan yazılımsal ve donanımsal sistemler kullanılması, derinlemesine savunma stratejisi zorunlu olması gerekmektedir.

10. MITM SALDIRILARINDAN KORUNMA YÖNTEMLERİ

MitM saldırılarının etkilerini önlemek için güçlü şifreleme protokolleri (ör. HTTPS, VPN), kimlik doğrulama yöntemleri (ör. MFA), güvenilir ağ kullanımı ve eğitimli kullanıcılar kritik öneme sahiptir.[22]

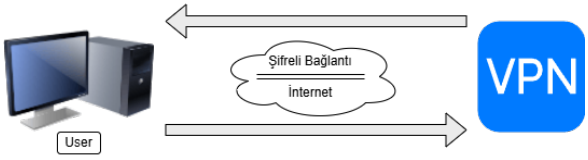
A. Şifreleme Kullanımı

Şifreleme, veri iletiminde gizliliği sağlamanın temel yöntemidir. TLS/SSL protokolleri bu amaca yönelik olarak yaygın şekilde kullanılır. Çoğu şifreleme protokolü, özellikle MitM saldırılarından korunmak için bir tür kimlik doğrulaması içermektedir. SSL/TLS sertifikaları, saldırganların güvenli bağlantılardan ödün vermek için fazladan adımlar atması gerektiğinden, etkin MitM saldırılarını da fazlasıyla zorlaştırır.

B. 10.2.VPN Kullanımı

VPN' ler, çevrimiçi veri trafiğini şifreleyerek kullanıcıları MitM saldırılarına karşı korur. Bir VPN kullanarak, ARP saldırılarının

çoğunu da büyük ölçüde engellenebilir. Özellikle herkese açık veya güvenilmeyen Wi-Fi ağlarında (gerekli olmadıkça) internete bağlandığımızda mutlaka bir VPN kullanmamız gerekmektedir. Çevrimiçi etkinliklerinizle ilgili hiçbir bilgiyi saklamayan veya paylaşmayan, sıkı bir kayıt tutmama politikasına sahip saygın VPN uygulamaları yapılacak saldırılarda bizi olabildiğince güvende tutacaktır.



Şekil 9. VPN kullanımı

C. *Güvenilir Ağlar Kullanma*

Bir ağın güvenilir olması, veri iletişimimizin saldırganlar tarafından ele geçirilme riskinin düşük olduğu anlamına gelir. Bilinmeyen veya halka açık Wi-Fi ağlarından uzak durmak, riskleri azaltabilir ve MitM saldırılarının en çok olarak gerçekleştiği yerlerin burası olduğu unutulmamalıdır. Bu tür saldırılarda saldırgan ağa katılıp trafiği izler, değiştirebilir ve yönlendirme yapabilir. Zorunlu kaldığımız durumlarda bağlanacağımız ağın güvenilir olduğunu kontrol etmeliyiz. Cihazlarımızda güvenlik duvarı ve güncel bir antivirüs uygulamaları bulundurmak önemlidir, kötü amaçlı girişimleri tespit edip engelleyebilir. Banka işlemleri ve hassas işlemler bu ağlardan giriş sağlanmamalıdır. Bağlantı ayarlarımızı kontrol etmekten vazgeçmeyelim. Bilmediğimiz ağlara otomatik bağlanmaktan çekinmeliyiz.

D. *Güncellemelerin Takibi*

Yazılım ve donanım güncellemeleri, bilinen güvenlik açıklarının kapatılmasını sağlar. Yazılımlar ve donanımlar zaman geçtikçe güvenlik açıklarına maruz kalabilir. Siber saldırganlar, bu açıkları kullanarak sistemlere izinsiz erişim sağlayabilir. Üreticiler, bu açıkları kapatmak ve sistemi güçlendirmek için düzenli olarak güncellemeler yayınlar. Güncellemelerin takip edilmemesi, bilinen zafiyetlere karşı savunmasız kalmamıza sebep oluşturabilir. Kullandığımız cihazlarda mutlaka otomatik güncelleştirmeleri aktif etmeliyiz. Modem yönlendirici cihazlarımızda (Akıllı TV, kamera, IoT cihazlar gibi) yazılım sürücülerini de kontrol edip göz ardı etmeden mutlaka güncellemeleri yapılmalıdır. Yapılan güncellemeler güvenlik dışında, performansı arttırır, iyileştirmeleri ve hata düzenlemelerini yapmaya olanak sağlar.[23]

E. *Sertifika Doğrulama*

Web sitelerinin sertifikalarını kontrol etmek, sahte sitelere yönlendirilme riskini azaltır. Sertifikalar, internet üzerindeki şifreli iletişim için kullanılan TLS/SSL protokollerinin bir parçasıdır. Web siteleri, kullanıcıların bağlantılarının güvenli olduğunu ve sahte bir siteye yönlendirilmediklerini göstermek için dijital sertifikalar kullanmaya olanak sağlar. Sahte veya geçersiz bir sertifika, bir Man-in-the-Middle (MitM) saldırısının önemli bir parçasıdır. Tarayıcıların adres çubuğunda bulunan kilit simgesi bağlantının şifreli olduğunu gösterir fakat kontrol sağlayıp sertifikanın geçerli olup olmadığını kontrol etmemiz gerekmektedir.[24]

11. BULGULAR

MitM Saldırılarının Yaygınlığı ve Etkisi ile MitM saldırıları, özellikle halka açık herkesin kullanabileceği ağlarda veri hırsızlığı ve saldırılar için yaygın olarak kullanılan bir yöntemdir. SSL/TLS protokollerinin etkili bir şekilde kullanılması, gerekli önlemlerin alınması bu saldırılara karşı önemli bir koruma sağlamaktadır. Bulut Bilişim Güvenlik Açıkları: Verilerin merkezi olarak depolanması, bulut bilişimde hem bir avantaj hem de bir risk yaratmaktadır. Özellikle zayıf şifreleme veya kimlik doğrulama sistemleri ciddi güvenlik açıklarına yol açabilir. Güvenlik önlemlerinin etkinliği şifreleme, VPN kullanımı ve kimlik doğrulama gibi yöntemlerin bir arada uygulanması, güvenlik tehditlerine karşı daha güçlü bir savunma mekanizması oluşturur.

12. TARTIŞMA

MitM Saldırılarında Farkındalık ve Eğitim: Kullanıcılar, genellikle MitM saldırılarının farkında olmadıkları için halka açık ağlarda güvenli bağlantılar kullanmaya dikkat etmemektedir. Bu, siber saldırganlar için fırsatlar yaratmaktadır. Eğitimin artırılması, bu saldırıların etkinliğini

azaltabilir. Bulut Bilişim ve Regülasyonlar: Veri gizliliği ve güvenliği konusunda uluslararası regülasyonların yetersizliği, bulut bilişimde güvenlik açıklarının artmasına neden olabilir. Regülasyonların ve standartların güçlendirilmesi, bulut ortamındaki güvenliği artırabilir. Siber güvenlik teknolojilerinin geleceği, yapay zekâ ve makine öğrenimi tabanlı güvenlik sistemleri hem MitM saldırılarını hem de bulut bilişim güvenlik açıklarını tespit etmekte etkili olabilir. Bu teknolojilerin daha yaygın kullanımı, güvenlik tehditlerini minimize etme potansiyeline sahiptir.

13. SONUÇ

Man-in-the-Middle (MitM) saldırıları ve bulut bilişimdeki güvenlik açıkları, dijital dünyada her geçen gün daha fazla önem kazanmaktadır. MitM saldırıları, bireylerin ve kurumların verilerini tehdit eden karmaşık yöntemler sunarken, bulut bilişim altyapılarında veri güvenliği ve sistem sürekliliği kritik birer konu olarak öne çıkmaktadır. Siber güvenlik önlemlerini artırmak, yalnızca teknolojik çözümlerle değil, aynı zamanda kullanıcı eğitimi ve farkındalığıyla da mümkündür.

MitM saldırıları karşısında güçlü şifreleme yöntemleri, VPN kullanımı ve güvenilir ağlara bağlı kalmak gibi stratejiler, bireysel ve kurumsal düzeyde etkili savunma mekanizmaları sunar. Öte yandan, bulut bilişimde güvenlik sağlamak için şifreleme, kimlik doğrulama, yedekleme ve düzenli güvenlik denetimleri gibi önlemler, veri gizliliği ve süreklilik açısından hayati önem taşır. Eğitim ve farkındalık çalışmaları ise insan faktöründen kaynaklanabilecek riskleri minimize eder.

Siber tehditlerin sürekli değiştiği bir dünyada, kullanıcılar ve kurumlar proaktif bir yaklaşım benimsemelidir. Teknolojik yeniliklerle birlikte riskler de artarken, siber güvenlik önlemlerini benimsemek yalnızca bir seçenek değil, bir zorunluluktur. Bu makale, MitM saldırıları ve bulut bilişimdeki güvenlik açıkları hakkında temel farkındalık oluşturmayı ve daha güvenli bir dijital ortamın oluşturulmasına katkıda bulunmayı hedeflemektedir.

14. TEŞEKKÜR

Bu çalışma Ankara Medipol Üniversite Siber Güvenlik Tezli Yüksek Lisans danışman hocam Dr. Öğr. Üye. Cevat ÖZARPA tarafından desteklenmiş olup teşekkürlerimi bir borç bilirim.

15. REFERENCES

[1] *E. Masum and R. Samet, "Mobil BOTNET İle DDOS Saldırısı," Bilişim Teknolojileri Dergisi, vol. 11, no. 2, pp. 111–121, Apr. 2018, doi: 10.17671/gazibtd.306612.*

[2] *N. Keskin, A. Nur Kıran, F. Kübra Eğdemir, and T. Eren, "Bulut Bilişim Güvenlik Gereksinimlerine Göre Çok Ölçütlü Karar Verme Yöntemleri İle Hizmet Sağlayıcı Seçimi." [Online]. Available: https:/dergipark.gov.tr/ubgmd*

[3] *I. Karabey Aksakallı, "Bulut Bilişimde Güvenlik Zafiyetleri, Tehditler ve Bu Tehditlere Yönelik Güvenlik Önerilerinin İncelenmesi."*

[4] *M. Z. Gündüz and R. Daş, "Kişisel Siber Güvenlik Yaklaşımlarının Değerlendirilmesi," DÜMF Mühendislik Dergisi, Jun. 2022, doi: 10.24012/dumf.1122997.*

[5] *"Bulut Bilişim Güvenliği."*

[6] *K. Yigit and F. Ayata, "Bulut Bilişim Mimarisi ve Güvenliği," Computer Science, Apr. 2024, doi: 10.53070/bbd.1448462.*

[7] *S. Hizal, Bulut Bilişim Güvenliği. 2023. [Online]. Available: www.nobelkitap.com*

[8] *C. Hatipoğlu and T. Tunacan, "Türkiye'de Siber Saldırı ve Tespit Yöntemleri: Bir Literatür Taraması," Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi, vol. 8, no. 1, pp. 430–445, Jun. 2021, doi: 10.35193/bseufbd.838732.*

[9] *A. Mallik, "Man in the Middle Attack: Understanding in Sımple Words," 2018.*

[10] *O. Fındık, "Şifrelemede Kaotik Sistemin Kullanılması".*

[11] G. Deliorman et al., "Bulut Sistem Güvenliği, Saldırılar ve Saldırılardan Korunmak için Geliştirilen Yöntemler."

[12] F. Shaar and A. Efe, "DDoS Attacks and Impacts on Various Cloud Computing Components."

[13] İ. Kara, "Kaba Kuvvet Saldırı Tespiti ve Teknik Analizi," Sakarya University Journal of Computer and Information Sciences, vol. 2, no. 2, pp. 61–69, Aug. 2019, doi: 10.35377/saucis.02.02.561844.

[14] Ş. Kara et al., "Ağ Sistemlerinin Güvenliği İçin Siber Saldırıların Ayrık Olaylı Sistem Tanımlama Tabanlı Modellenmesi ve Simülasyonu," 2023.

[15] M. M. Şimşek and E. Atılgan, "Attacks on Availability of IoT Middleware Protocols: A Case Study on MQTT," Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi, vol. 4, no. 2, pp. 16–27, Jun. 2023, doi: 10.53608/estudambilisim.1297052.

[16] S. Z. Dicle, "Man-In-The-Middle Attack," European Journal of Science and Technology, Oct. 2022, doi: 10.31590/ejosat.1187984.

[17] E. Resoğlu, B. Mühendisliği, A. Dalı, and B. T. Programı, "Kablosuz Ağlarda Güvenlik," 2008.

[18] Z. Čekerevac, Z. Dvorak, L. Prigoda, and P. Čekerevac, "Internet of Things and The Manin the Middle Attacks- Security and Economıc Risks," MEST Journal, vol. 5, no. 2, pp. 15–5, Jul. 2017, doi: 10.12709/mest.05.05.02.03.

[19] H. Toprak and S. Kardas, "Layer 2 Internal Network Attacks And The Countermeasures," 2020. [Online]. Available: https://www.researchgate.net/publication/348151619

[20] A. Çoroz and E. Akyüz, "Türk Mühendislik Araştırma ve Eğitimi Dergisi IoT and Cloud Based Application for Smart Cities Akıllı Şehirler için Nesnelerin İnterneti ve Bulut Tabanlı bir Uygulama," 2022.

[21] A. Koçhan, "AKILLI KENTLER VE DİJİTAL (SİBER) GÜVENLİK * Hayriye ŞENGÜN 1** Seçil Gül MEYDAN YILDIZ ****." [Online]. Available: https://orcid.org/0000-0001-9869-4159

[22] İ. AVCI, "Akıllı ulaşım sistemlerinde siber saldırılar ve önlemler," Akıllı Ulaşım Sistemleri ve Uygulamaları Dergisi, vol. 6, no. 1, pp. 194–208, Mar. 2023, doi: 10.51513/jitsa.1224909.

[23] I. Karabey Aksakallı, "Bulut Bilişimde Güvenlik Zafiyetleri, Tehditler ve Bu Tehditlere Yönelik Güvenlik Önerilerinin İncelenmesi."

[24] A. Efe, G. Kalkanci, M. Donk, S. Cihangir, and Z. Uysal, "A Hidden Hazard: Man-in-The-Middle Attack in Networks."

[25] F. Shaar and A. Efe, "DDoS Attacks and Impacts on Various Cloud Computing Components."

ЖАСАЛМА ИНТЕЛЛЕКТ АРКЫЛУУ КЫРГЫЗ ТИЛИНДЕГИ САНДЫК ФРАЗЕОЛОГИЗМДЕРДИ КОТОРУУДАГЫ МАДАНИЙ-СЕМАНТИКАЛЫК АДЕКВАТТУУЛУК МАСЕЛЕЛЕРИ

Бирназарова Айнура Жапаралиевна
Национальная Академия Наук КР, Кыргызстан, Бишкек birnazar.ripk22@gmail.com
Бексултанова Гулзада Азимкановна
Международный университет в Кыргызстане, Кыргызстан, Бишкек gulsada@bk.ru

PROBLEMS OF CULTURAL AND SEMANTIC ADEQUACY IN TRANSLATING NUMERICAL PHRASEOLOGICAL UNITS IN THE KYRGYZ LANGUAGE VIA ARTIFICIAL INTELLIGENCE

Birnazarova Ainura Zhaparalievna
National Academy of Sciences of the Kyrgyz Republic, Kyrgyzstan, Bishkek
birnazar.ripk22@gmail.com
Beksultanova Gulzada Azimkanovna
International University of Kyrgyzstan, Kyrgyzstan, Bishkek
gulsada@bk.ru

Аннотация

Макалада заманбап компьютердик лингвистиканын жана жасалма интеллекттин (ЖИ) эң актуалдуу маселелеринин бири - ресурсу аз тилдердин маданий-семантикалык өзгөчөлүктөрүн машиналык котормодо жана чоң тилдик моделдерде (LLM) трансформациялоо көйгөйү кыргыз тилинин мисалында изилденди.

Изилдөөнүн объектиси катары улуттук дүйнө таанымдын, мифологиянын жана философиялык концепттердин негизин түзгөн сан компоненттүү (3, 7, 9, 40 ж.б.) кыргыз фразеологизмдери тандалып алынды. Эксперименталдык бөлүмдө заманбап ЖИ моделдеринин (ChatGPT, Google Gemini) жана нейрондук машиналык котормо системаларынын (Google Translate, DeepL) сандык фразеологизмдерди англис жана орус тилдерине которуудагы кудурети, контекстти кабылдоосу жана концептуалдык каталары салыштырмалуу талдоого алынды.

Изилдөөдө ЖИ моделдеринин маданий идиомаларды көбүнчө сөзмө-сөз которору же ортомчу тилдин таасиринен улам семантикалык бурмалоолорго жол берери аныкталды. Макаланын жыйынтыгында ресурсу аз тилдердин маданий кодун санариптик мейкиндикте сактоо жана ЖИ моделдерин лингвокультурологиялык параллелдүү корпустар менен тереңдетип машыктыруу (*fine-tuning*) боюнча практикалык сунуштар берилет.

Ачкыч сөздөр: *жасалма интеллект, чоң тилдик моделдер (LLM), сандык фразеологизмдер, маданий-семантикалык адекваттуулук, ресурсу аз тилдер, машиналык котормо, лингвокультурология, семантикалык бурмалануу.*

Abstract

The article investigates one of the most pressing issues in modern computational linguistics and artificial intelligence (AI): the problem of translating the cultural and semantic features of low-resource languages within machine translation and large language models (LLMs), using the Kyrgyz language as a case study. Kyrgyz numeral idioms (containing components such as 3, 7, 9, 40, etc.), which form the foundation of the national worldview, mythology, and philosophical concepts, were chosen as the object of this study.

In the experimental section, a comparative analysis was conducted to evaluate the capabilities, context comprehension, and conceptual errors of modern AI models (ChatGPT, Google Gemini) and neural machine translation systems (Google Translate, DeepL) when translating these numeral idioms into English and Russian. The

study revealed that AI models often translate cultural idioms literally or generate semantic distortions due to the influence of a pivot language. Consequently, the article provides practical recommendations for preserving the cultural code of low-resource languages in the digital space and fine-tuning AI models using linguocultural parallel corpora.

Keywords: artificial intelligence, large language models (LLMs), numeral idioms, cultural-semantic adequacy, low-resource languages, machine translation, linguoculture, semantic drift.

КИРИШҮҮ

Глобализация жана коомду жапырт санариптештирүү заманында табигый тилди иштетүү (NLP – Natural Language Processing) жана жасалма интеллект (ЖИ) технологиялары тил аралык байланыштын негизги куралына айланды. Чоң тилдик моделдердин (LLM- Large Language Models) жана нейрондук машиналык котормо системаларынын болуп көрбөгөндөй өнүгүшү маалымат алмашууну тездеткени менен, тилдердин ортосундагы санариптик теңсиздик маселесин да курчутту.

Дүйнөлүк лингвистикалык технологияларда кыргыз тили сыяктуу санариптик ресурсу чектелген тилдер "ресурсу аз тилдер" деп аталып, алардын маданий, көркөм жана контексттик өзгөчөлүктөрүн жасалма интеллектке туура үйрөтүү бүгүнкү күндүн эң татаал илимий-практикалык көйгөйлөрүнүн бири бойдон калууда (Жоробекова, Исаева, 2023).

Тилдин эң жогорку маданий жана философиялык катмары анын фразеологиялык фондунда сакталат. Өзгөчө кыргыз тилиндеги сан компоненттүү фразеологизмдер (мисалы: жети атаны билүү, бир жеңден кол, бир жакадан баш чыгаруу, тогуз жолдун тоому, үч уктаса түшкө кирбеген, беш колдой билүү, кыздын кырк чачы улуу ж.б.) жөн гана математикалык бирдиктерди туюнтпостон, улуттук мифологиянын, көчмөнчүлүк философиясынын жана маданий коддун үлгүсү болуп саналат. Бул сандар улуттук аң-сезимде метафоралык жана символикалык терең мааниге ээ. Мисалы, "тогуз" саны кыргыз контекстинде көптүктү, чексиздикти же маанилүү стратегиялык түйүндү билдирсе, "жети" саны ыйыктыкты жана ырааттуулукту туюнтат.

Бүгүнкү күндө Google Translate, Yandex Translate же ChatGPT сыяктуу системалар түз маанидеги тексттерди жана стандарттуу сүйлөмдөрдү жогорку деңгээлде которо алганы менен, тилдин метафоралык катмарына, тактап айтканда, фразеологизмдерге келгенде олуттуу кемчиликтерди көрсөтүүдө. Машиналык котормодогу эң чоң ката - бул фразеологиялык бирдиктерди сөзмө-сөз которуу болуп саналат (Мамбеталиева, 2022), мунун кесепетинен тексттин түпкү маданий мааниси толугу менен жоголот. Мындан сырткары, ЖИ моделдери кыргыз тилинен англис же башка батыш тилдерине которууда көбүнчө орус тилин ортомчу тил катары колдонушат. Бул эки баскычтуу трансформация процессинде маданий адекваттуулук деңгээли төмөндөп, семантикалык бурмалануу феномени келип чыгат.

Изилдөөнүн максаты - заманбап жасалма интеллект моделдеридин кыргыз тилиндеги сандык фразеологизмдердин маданий-семантикалык адекваттуулугун сактоо менен башка чет тилдерге которуу мүмкүнчүлүктөрүн жана кетирген концептуалдык каталарын илимий-эксперименталдык өңүттөн изилдөө, ошондой эле аларды чечүүнүн лингво-технологиялык

жолдорун сунуштоо болуп саналат. Бул маселени изилдөө лингвистика менен маалыматтык технологиялардын (IT) кесилишинде жаңы илимий багыттарды ачууга жана кыргыз тилинин санариптик мейкиндикте өзүнүн улуттук өзгөчөлүгүн жоготпой, сабаттуу интеграцияланышына шарт түзөт.

ТЕОРИЯЛЫК жана ПРАКТИКАЛЫК НЕГИЗДЕР

Тилдин курамындагы фразеологиялык бирдиктер - бул жөн гана сөз айкаштары эмес, ошол тилде сүйлөгөн элдин кылымдар бою топтогон тажрыйбасынын, тарыхынын, психологиясынын жана улуттук дүйнө таанымынын концентратталган көрүнүшү болуп саналат. Лингвокультурология илиминде мындай бирдиктер "маданий коддор" деп аталат.

Маданий коддор курчап турган дүйнөнү кабылдоонун улуттук өзгөчөлүгүн калыптандырат жана бир тилден экинчи тилге механикалык түрдө которууга мүмкүн болбогон лингвистикалык феномендерди жаратат (Маслова, 2018). Бул контекстте кыргыз тилиндеги сан компоненттүү (нумерологиялык) фразеологизмдер өзгөчө орунду ээлейт, анткени алардын курамындагы сандар өздөрүнүн алгачкы математикалык (дискреттик) маанисинен толугу менен ажырап, метафоралык, символикалык жана концептуалдык деңгээлге көтөрүлгөн.

Кыргыз улуттук аң-сезиминде 3, 7, 9, 40 сыяктуу сандар көчмөнчүлүк турмуш-тиричилигине, космогониялык түшүнүктөрүнө жана мифологиялык ишенимдерине түздөн-түз байланышкан. Мисалы, кыргыз дүйнө таанымында "жети" саны - ыйыктыктын, системалуулуктун жана адамдын коомдогу ордун аныктоонун символу. "Жети атасын билүү" концепциясы жөн гана жети муундун тизмеси эмес, ал генеалогиялык эс тутумду, ген тазалыгын жана социалдык жоопкерчиликти сактоонун институту болуп саналат. Ал эми "жети өлчөп, бир кес" фразеологизминдеги "жети" саны конкреттүү математикалык аракетти эмес, чечим кабыл алуудагы максималдуу этияттыкты, акылмандыкты жана ички ырааттуулукту туюнтат (Ахматова, 2021). Ошо сыяктуу эле, "тогуз" саны кыргыз маданиятында чексиздиктин, толуктуктун жана жогорку баалуулуктун метафорасы катары чыгат. Тилдеги "тогуз жолдун тоому" деген туруктуу айкаш геометриялык тогуз жолдун кесилишин билдирбейт. Ал - маанилүү стратегиялык борборду, тагдыр чечүүчү чекитти мүнөздөгөн татаал лингвокультурологиялык концепт. Бул жердеги "тогуз" - көптүк маанинин эң жогорку чеги. Демек, сандык фразеологизмдердин мааниси анын курамындагы компоненттердин маанилеринин суммасына барабар эмес. Фразеологизмдин мааниси анын ички маданий контекстинен гана жаралат.

Жасалма интеллект (ЖИ) жана чоң тилдик моделдер (LLM) үчүн дал ушул жерден эң чоң лингвистикалык тоскоолдук келип чыгат. Компьютердик лингвистикада маданий-семантикалык адекваттуулук - бул лексикалык бирдиктерди сөздүктөгү маанисине карап синонимдөө эмес, баштапкы тилдеги маданий концептти кабыл алуучу тилдин маданий мейкиндигине контексттик жана стилистикалык жоготуусуз трансформациялоо болуп саналат (Wierzbicka, 2020). Нейрондук тармактар текстти иштетүүдө математикалык вектрлерге жана сөздөрдүн бирге кездешүү

ыктымалдуулугуна таянгандыктан, алар фразеологизмдин артында турган улуттук мифологияны жана коннотациялык (өтмө) маанини түз кабылдай алышпайт (Koehn, 2020). Андан сырткары, ресурсу аз тилдерди которууда ЖИ моделдери көп учурда ортомчу тил феноменине туш болушат. Башкача айтканде, кыргызча фразеологизм адегенде орус тилинин структурасына көчүрүлүп, андан кийин англис тилине которулат. Бул эки баскычтуу трансформация процессинде баштапкы метафоранын мааниси эки эселенген семантикалык чыпкадан өтүп, жыйынтыгында семантикалык бурмалануруга дуушар болот (Aitbaev, Tukeev, 2023). Натыйжада, машина тексттин стилин, улуттук боёгун (колоритин) жоготуп, окурманда туура эмес ассоциацияларды жараткан жасалма структураны сунуштайт. Ошентип, сандык фразеологизмдерди ЖИ аркылуу которуу маселеси - жөн гана техникалык мүчүлүштүк эмес, бул тил илиминин тереңиндеги семантикалык кубулуштар менен жасалма интеллекттин алгоритмдик чектелүүсүнүн ортосундагы концептуалдык карама-каршылык болуп саналат.

Жасалма интеллект моделдеринин жана нейрондук машиналык котормо системаларынын кыргыз тилиндеги сандык фразеологизмдерди кабылдоо мүмкүнчүлүктөрүн иш жүзүндө текшерүү максатында атайын лингвистикалык сыноо (эксперимент) жүргүзүлдү. Эксперимент үчүн бүгүнкү күндүн эң популярдуу жана кеңири колдонулган үч платформасы тандалып алынды: нейрондук машиналык котормо системасы Google Translate жана ири алдыңкы чоң тилдик моделдер (LLM) - ChatGPT (GPT-4o версиясы) жана Google Gemini. Сыноо процессинде моделдерге кыргыз тилинин контексттик жана маданий өзгөчөлүгүн камтыган, курамында сандык компоненттери бар туруктуу сөз айкаштарын англис жана орус тилдерине түз которуу тапшырмасы берилди. Төмөндө тандалып алынган негизги лингвокультурологиялык бирдиктердин ЖИ моделдери тарабынан которулуу өзгөчөлүктөрүнө деталдуу семантикалык талдоо берилет.

«*Жети өлчөп, бир кес»* фразеологизминин котормосун талдоо

Бул фразеологизм терең ойлонууну жана этияттыкты билдирет. Батыш маданиятында мунун түз математикалык эквиваленти ("эки жолу өлчөө") же метафоралык эквиваленти ("секирүүдөн мурун алдыга кара") бар.

Google Translate: Бул система фразеологизмди англис тилине "Measure seven times, cut once", ал эми орус тилине "Семь раз отмерь, один раз отрежь" деп которгон. Орус тилиндеги котормосу маданий жана маанилик жактан толук адекваттуу, анткени орус маданиятында да дал ушул сандык идиома колдонулат. Бирок англис тилине которууда ЖИ кыргызча "жети" санын механикалык түрдө сактап калган. Англис тилдүү чөйрөдө бул идиома маданий код катары калыптанбагандыктан, түз маанисинде кабыл алынат.

ChatGPT жана Google Gemini: Чоң тилдик моделдер өзүлөрүнүн лингвистик базасынын кеңдигинен улам, бул айкаштын маанисин түшүндүрүүгө аракет кылышкан. ChatGPT контекстти эске алып, фразеологизмдин англисче нукура маданий эквиваленти катары "Measure twice, cut once"

же "Look before you leap" варианттарын сунуштай алган. Бул моделдердин семантикалык деңгээлде иштеп жатканын көрсөтөт.

«Үч уктаса түшкө кирбеген» фразеологизминин котормосун талдоо. Бул бирдик такыр күтүлбөгөн, ишенгис же адамдын акылына сыйбаган окуяларды мүнөздөө үчүн колдонулат. "Үч" саны бул жерде чектүүлүктүн жана убакыттык мейкиндиктин метафорасы кызматын аткарат.

Google Translate: Система англис тилине "Not dreamed of even in three sleeps", орус тилине "Даже во сне трижды не снилось" деп сөзмө-сөз которгон. Натыйжада англис тилинде таптакыр маанисиз, стилистикалык жактан ката жана кабыл алуучу маданият үчүн чоочун фраза пайда болгон. Орус тилиндеги вариант да жасалма угулат, анткени орус тилинин жандуу фразеологиясында мындай түзүм жок.

ChatGPT жана Google Gemini: Эки модель тең фразеологизмди сөзмө-сөз которуудан качып, анын семантикалык ядросун кармап калууга жетишкен. Алар англис тилиндеги маданий эквиваленттерди - "Never in a million years" же "Beyond one's wildest dreams" варианттарын сунушташкан. Бул ЖИнин идиомалык деңгээлде тилдик трансформацияны адекваттуу аткара аларын далилдейт, бирок ал үчүн моделге стилистикалык тууралоо талап кылынат.

«Тогуз жолдун тоому» фразеологизминин котормосун талдоо. Кыргыздын дүйнө таанымында эң татаал концепттердин бири. "тогуз" саны - көптүктүн, татаалдыктын жана чексиздиктин метафорасы.

Google Translate: Англисчеге "The intersection of nine roads", орусчага "Перекресток девяти дорог" деп гана берген. Бул механикалык-геометриялык котормо фразеологизмдин артында турган "стратегиялык маанилүү борбор, тагдыр чечүүчү чекит" деген терең коннотациялык маанини толугу менен жок кылат.

ChatGPT жана Google Gemini: Бул моделдер да "тогуз" санын түз маанисинде кабыл алып, ката кетиришкен. Алар текстти "Crossroads of nine ways" деп которуп, андан кийин гана кашанын ичинде бул жерде "чоң кесилиш же маанилүү борбор" тууралуу сөз болуп жатканын түшүндүрүп жазышкан. Демек, ЖИ текстти өзүнөн маданий кодду дароо тааный алган эмес, аны кошумча түшүндүрмө аркылуу гана чечмелеген.

Жалпысынан эксперимент көрсөткөндөй, жөнөкөй машиналык котормо системалары (Google Translate) фразеологизмдерди түз лексикалык бирдик катары кабыл алып, сөзмө-сөз которууга ыкташат. Ал чоң тилдик моделдер (ChatGPT, Gemini) контекстти жана идиомалык маанини кыйла тереңирээк талдай алышат, бирок кыргыз тили сыяктуу ресурсу аз тилдердин өзгөчө сандык метафораларына келгенде, алар дагы деле маданий кодду толук чечмелей албай, ортомчу тилдердин калыбына салып алышат.

ТАЛКУУ

Жүргүзүлгөн лингвистикалык эксперименттин жыйынтыгы көрсөткөндөй, жасалма интеллект системалары кыргыз тилиндеги сан компоненттүү фразеологизмдерди которууда бир катар концептуалдык жана структуралык каталарды кетирет. Бул мүчүлүштүктөр жөнөкөй эле техникалык мүчүлүштүк эмес, алардын артында терең лингвистикалык, маданий жана алгоритмдик себептер жатат. ЖИ моделдеринин маданий адекваттуулукка жетише албаган негизги көйгөйлөрүн төмөнкүдөй классификациялоого болот:

-*контекстти жана маданий коннотацияны кабыл албагандык* (сөзмө-сөз которуу).

Жасалма интеллекттин эң негизги кемчилиги - текстти денотативдик (түз) деңгээлде гана кабылдап, коннотативдик (өтмө, жашыруун) маанини өткөрүп жибергендигинде. Кыргыз тилиндеги сандык фразеологизмдердин семантикасы алардын сырткы лексикалык курамына эмес, ички маданий контекстине көз каранды. ЖИ модели идиоманы кездештиргенде, анын архитектурасы алгач сөздөрдүн өзүнчө маанилерин издейт. Натыйжада, улуттук аң-сезимдеги философиялык метафора болгон сандар кадимки математикалык эсеп бирдиги катары кабыл алынып, сөзмө-сөз которулуп кетет. Бул процесс максаттуу тилдин (кабыл алуучу тилдин) маданий мейкиндигинде баштапкы маанинин толук жоголушуна алып келет.

-санариптик тексттик корпустардын жетишсиздиги (low-resource language феномени).

Чоң тилдик моделдердин (LLM) акылдуулугу жана тилди сезүү кудурети түздөн-түз алар машыккан маалыматтардын (тексттик корпустардын) көлөмүнө байланыштуу. Англис, орус же кытай тилдери санариптик дүйнөдө миллиондогон параллелдүү сөздүктөр, көркөм адабияттар жана илимий эмгектер менен камсыз кылынгандыктан, ЖИ бул тилдердин идиомаларын абдан так тааныйт. Ал эми кыргыз тили санариптик ресурсу аз тилдердин катарына киргендиктен, ачык интернет мейкиндигинде кыргызча-англисче сапаттуу фразеологиялык параллелдүү корпустар дээрлик жокко эсе (Turdaliev, Sultanov, 2024). Моделдердин колунда жетиштүү маалымат (датасет) болбогондуктан, алар фразеологизмдин туура эквивалентин алгоритмдик түрдө эсептеп чыга албай жатышат.

-*ортомчу тилдин терс таасири.* Компьютердик лингвистикада кыргыз тилинен англис тилине же тескерисинче которуу түз ишке ашпайт. Көпчүлүк ЖИ моделдери жана машиналык котормо системалары кыргызча текстти адегенде орус тилине которуп, андан кийин гана орус тилинен англис тилине трансформациялоочу көп баскычтуу моделди колдонушат. Бул баскычта эки эселенген семантикалык чыпкалоо жүрөт.

Кыргыз тилиинин маданий коду алгач орус тилинин структурасына ылайыкташып, андан кийин англис тилинин калыбына куюлат. Мындай механикалык трансфердин натыйжасында семантикалык бурмалануу кубулушу келип чыгат - текст өзүнүн улуттук боёгун, стилистикалык өзгөчөлүгүн жана эмоционалдык экспрессиясын толугу менен жоготот.

-статистикалык ыктымалдыктын маданий ийкемдүүлүктөн үстөмдүгү. Заманбап жасалма интеллект нейрондук тармактардын негизинде иштейт жана кийинки сөздүн кандай болорун статистикалык ыктымалдык менен гана болжолдойт (Vaswani et al., 2017). Ал фразеологизмдин артында турган улуттук мифологияны, тарыхый шарттарды жана көчмөн улуттун психологиясын "сезе" албайт. Машина үчүн "тогуз жол" - бул жөн гана тогуз даана жолдун математикалык айкалышы. Моделдердин мындай архитектуралык чектелүүсү тилдин жандуу, ийкемдүү жана образдуу табияты менен карама-каршы келип, санариптик котормодогу эң чоң тоскоолдук бойдон калууда.

ЧЕЧҮҮ ЖОЛДОРУ ЖАНА СУНУШТАР

Жасалма интеллект системаларында кыргыз тилиндеги сандык фразеологизмдерди которууда келип чыккан маданий-семантикалык адекваттуулук маселелерин чечүү бир гана техникалык же бир гана лингвистикалык мүнөздөгү кадамдар менен чектелбейт. Бул көйгөйдү жоюу үчүн тил илими менен маалыматтык технологиялардын (IT) кесилишинде комплекстүү, системалуу жана узак мөөнөттүү инновациялык мамиле талап кылынат. Кыргыз тилинин улуттук-маданий кодун санариптик мейкиндикте сактоо жана ЖИ моделдеринин котормо сапатын жогорулатуу үчүн төмөнкүдөй практикалык кадамдарды ишке ашыруу зарыл:

-параллелдүү санариптик фразеологиялык корпустарды түзүү.

Бүгүнкү күндө кыргыз тилиндеги машиналык котормонун сапатын көтөрүүнүн эң башкы шарты - бул сапаттуу санариптик маалыматтар базасын (датасеттерди) түзүү болуп саналат. Тилчи-лингвисттер атайын сандык жана башка маданий катмары терең фразеологизмдердин контексттик параллелдүү корпустарын (мисалы: кыргызча-англисче же французча., кыргызча-орусча, кыргызча-түркчө ж.б) иштеп чыгышы керек. Бул корпустарда фразеологизмдер жөн гана сөзмө-сөз которулбастан, алардын түз мааниси, өтмө мааниси, маданий контексти жана максаттуу тилдеги идеалдуу идиомалык эквиваленттери кошо камтылууга тийиш. Мындай адистештирилген корпустар жасалма интеллектти алгоритмдери үчүн негизги лингвистикалык багыт берүүчү курал болуп калат.

-моделдерди адистештирип машыктыруу (Fine-tuning)

Даярдалган параллелдүү фразеологиялык корпустардын негизинде ачык булактуу чоң тилдик моделдерди (мисалы: Llama 3, Mistral же ата мекендик ЖИ долбоорлорун) кыргыз тилинин маданий өзгөчөлүктөрүнө ылайыктап тереңдетип машыктыруу зарыл (Zong et al., 2021). Бул процесс моделдин жалпыланган статистикалык ыктымалдыктардан баш тартып, кыргыз тилинин уникалдуу лингвокультурологиялык эрежелерин өздөштүрүүсүнө жол ачат. Натыйжада, ЖИ тексти иштетүүдө "жети өлчөп, бир кес" деген айкашты көргөндө, аны механикалык түрдө сөздөргө бөлбөстөн, бирдиктүү маданий концепт катары таанып, дароо кабыл алуучу тилдин метафоралык эквивалентине багыттай алат.

-промт-инжиниринг (Prompt Engineering) стратегияларын өнүктүрүү

Учурдагы ЖИ моделдери менен иштөөдө маданий адекваттуулукка жетишүү үчүн колдонуучулардын промт (тапшырма) берүү маданиятын жогорулатуу маанилүү. Моделге түз которуу тапшырмасын бергендин ордуна, контекстти, стилди жана маданий эквивалент табуу зарылчылыгын түшүндүргөн көп баскычтуу промтторду колдонуу сунушталат. Мисалы, ЖИге: "Бул кыргыз тилиндеги сандык фразеологизм, муну сөзмө-сөз которбостон, анын улуттук метафоралык маанисин сактап, француз тилиндеги маданий аналогун тап" деген сыяктуу так чектөөлөр жана мисалдар менен кайрылуу котормонун сапатын дароо бир нече эсеге жогорулатат.

-ортомчу тилсиз (Direct Translation Model) түз которуу моделдерин иштеп чыгуу

Орус тилинин ортомчу курал катары таасирин азайтуу үчүн түз багыттагы санариптик көпүрөлөрдү куруу керек. Нейрондук машиналык котормо системаларын түздөн-түз тилдик багыттар боюнча машыктыруучу трансфердик технологияларды өнүктүрүү зарыл. Бул семантикалык бурмалануунун алдын алып, маалыматтын эки баскычтуу чыпкадан өтүүдөгү маданий жоготууларын минималдаштырат.

-компьютердик лингвисттердин жана IT адистеринин биргелешкен ишмердүүлүгү.

Бул маселени чечүүдөгү эң маанилүү кадам - тармактар аралык синергияны түзүү. Тил илиминин теориялык негиздери (лингвокультурология, семантика, фразеология) маалыматтык технологиялардын (машиналык үйрөтүү, NLP, нейрондук тармактар) практикалык куралдары менен биригиши керек. Жогорку окуу жайларында жана илимий борборлордо компьютердик лингвистика лабораторияларын түзүү, лингвисттер менен программисттердин биргелешкен "Алтын көпүрө" сыяктуу санариптик долбоорлорун мамлекеттик жана гранттык деңгээлде колдоо - кыргыз тилинин санариптик келечегин камсыз кылуунун негизги кепилдиги болуп саналат.

КОРУТУНДУ

Заманбап жасалма интеллекттин жана чоң тилдик моделдердин (LLM) болуп көрбөгөндөй өнүгүшү адамзаттын тил аралык байланышын жаңы деңгээлге көтөргөнү менен, ресурсу аз тилдердин маданий-семантикалык өзгөчөлүктөрүн санариптик мейкиндикте туура чагылдыруу маселесин күн тартибине койду. Бул изилдөөнүн алкагында кыргыз тилиндеги сан компоненттүү фразеологизмдердин жасалма интеллект (ЖИ) моделдери тарабынан которулушун маданий-семантикалык адекваттуулук өңүтүнөн талдоо бир катар маанилүү илимий жана практикалык жыйынтыктарды чыгарууга мүмкүндүк берди. Биринчиден, кыргыз тилиндеги сандык фразеологизмдер жөнөкөй эле математикалык бирдиктер эмес, алар кылымдар бою калыптанган улуттук маданий коддор, мифологиялык жана көчмөнчүлүк философиясынын концентратталган көрүнүшү болуп саналат. Бул бирдиктердин семантикасы анын курамындагы өзүнчө сөздөрдүн маанилеринин суммасынан алда канча кеңири жана ички маданий контекстке гана байланган. Экинчиден, жүргүзүлгөн лингвистикалык эксперимент көрсөткөндөй, заманбап ЖИ системалары

(айрыкча Google Translate сыяктуу нейрондук машиналык котормолор) фразеологизмдерди түз лексикалык деңгээлде кабылдап, сөзмө-сөз которууга ыкташат. Натыйжада, кабыл алуучу тилдин маданий мейкиндигинде баштапкы метафоралык маани толугу менен жоголот. Ал эми ChatGPT жана Google Gemini сыяктуу чоң тилдик моделдер идиомалык маанини кыйла тереңирээк талдай алышканы менен, кыргыз тилинин санариптик базасынын аздыгынан жана ортомчу тилдин таасиринен улам семантикалык бурмалануларга жол беришет. Үчүнчүдөн, бул маселени чечүү бир гана техникалык алгоритмдерди жаңылоо менен чектелбейт. Кыргыз тилинин маданий кодун ЖИге туура үйрөтүү үчүн адистештирилген санариптик параллелдүү фразеологиялык корпустарды (датасеттерди) түзүү, моделдерди максаттуу түрдө машыктыруу ($fine-tuning$) жана промт-инжиниринг стратегияларын өнүктүрүү зарыл.

Жыйынтыктап айтканда, жасалма интеллект келечекте кыргыз тилинин көркөмдүгүн, эмоционалдык боёгун жана терең маданий катмарын жеткире алган деңгээлге жетиши үчүн тил илими (лингвокультурология, семантика) менен IT тармагынын биргелешкен синергиясы аба менен суудай зарыл. Лингвисттер менен программисттердин тармактар аралык биргелешкен изилдөөлөрү гана кыргыз тилинин глобалдык санариптик экосистемада өзүнүн улуттук өзгөчөлүгүн жана суверенитетин жоготпой, сабаттуу өнүгүшүнө жол ачат.

Колдонулган адабияттардын тизмеси:

Ахматова, А. Т. (2021). *Кыргыз тилиндеги сан компоненттүү фразеологизмдердин лингвокультурологиялык табияты*. Бишкек: Илим.

Жоробекова, А. А., & Исаева, Г. К. (2023). Компьютердик лингвистика жана ресурсу аз тилдерди санариптештирүү маселелери. *КМУнун Жарчысы*, (2), 45-51.

Мамбеталиева, N. S. (2022). Машиналык котормодогу семантикалык трансформациялар жана адекваттуулук көйгөйү. *Тил жана маданият*, 14(3), 112-119.

Wierzbicka, A. (2020). *Semantics, Culture, and Cognition: Universal Human Concepts in Culture-Specific Configurations*. Oxford University Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press.

Turdaliev, K., & Sultanov, B. (2024). Building Parallel Corpora for Central Asian Languages: Challenges and Solutions in Kyrgyz-English Machine Translation. *Proceedings of the Association for Computational Linguistics (ACL)*, 324-331.

Маслова, В. А. (2018). *Лингвокультурология: Учебное пособие для студентов высших учебных заведений*. Москва: Академия.

Zong, C., Xia, R., & Zhang, J. (2021). *Text Data Mining*. Springer, Singapore.

Aitbaev, R., & Tukeev, U. (2023). Semantic Drift and Pivot Language Effects in Low-Resource Turkic Languages Translation. *International Journal of Computer Processing of Languages*, 29(4), 501-515.

Türk Dünyası Kritik Enerji Altyapılarında Yapay Zekâ Destekli Siber Tehdit Değerlendirmesi: Karşılaştırmalı Bir Çerçeve Önerisi

Taha Tokgöz
Yönetim Bilişim Sistemleri Bölümü
Ankara Medipol Üniversitesi
Ankara, Türkiye
taha@tahatokgoz.com

Cevat Özarpa
Siber Güvenlik Anabilim Dalı
Ankara Medipol Üniversitesi
Ankara, Türkiye
cevat.ozarpa@ankaramedipol.edu.tr

***Özet*—*Bu araştırmanın temel amacı, Türkiye, Kazakistan, Azerbaycan, Türkmenistan ve Özbekistan'da enerji altyapısında yer alan SCADA ve ICS sistemlerinin siber güvenlik gelişmişlik durumunu kıyaslamak, buna ek olarak yapay zekâ destekli bir risk değerlendirme yaklaşımı sunmaktır. Farklılık yaratan yönü ise, mevcut çalışmalarda eksik olan bölgesel boyutta karşılaştırma sunmasıdır. Beş ülkede faaliyet gösteren enerji kuruluşlarının kullandığı kontrol sistemleri, hem IEC 62443 hem de NIST CSF ölçütleriyle incelenmiştir. Protokol düzeyinde Modbus/TCP, DNP3 ile OPC UA üzerinde görülen açıklar, MITRE ATT&CK for ICS sınıflamasına göre eşleştirilerek değerlendirilmiş. Literatürde özellikle Türk devletlerinin benzer altyapılara sahip olduğu gerçeği dikkate alınarak yapılan ortak tehdit analizleri oldukça sınırlıdır. Bu boşluğu doldurabilmek adına açık erişimli veri kaynaklarından faydalanılmış: örnek olarak MITRE ATT&CK for ICS, CISA'nın yayınladığı uyarılar ve Ulusal Zafiyet Veritabanı. Ülkeler için ayrı ayrı oluşturulan saldırı modelleri, karşılaştırmalı biçimde çözümlenmiş. Değerlendirme sürecinde IEC standardının tanımladığı SL1 ila SL4 arası güvenlik katmanlarına odaklanılmış. Son aşamada basitleştirilmiş, makine öğrenmesine dayalı bir skorlama yöntemi tasarlanmış. Bazı Türk Dünyası ülkelerinde eski SCADA sistemleri hâlâ kullanılıyor; bu da genel savunma yapılarını zayıflatıyor. Ağların parçalanmamış olması yaygın bir sorun gibi görünüyor. Kimlik kontrolünde boşluklar dikkat çeken noktalardan biri. Ülke bazlı dijital hazır bulunuşluk farklılıkları net şekilde gözlemlenebiliyor. Yapay zekânın anormallikleri algılama potansiyeli değerlendirildiğinde, bölgede nasıl işleyebileceği daha iyi anlaşılıyor. Bu araştırma, küresel değil bölgesel ölçekte ilk kez karşılaştırılabilir veriler üretiyor. Sonuç olarak, TDT içinde benzer altyapı güvenliği stratejilerinin şekillendirilmesi için sağlam bir başlangıç noktası oluşturuluyor.*

***Anahtar Kelimeler*—**Kritik Altyapı Güvenliği, SCADA/ICS, Yapay Zekâ Destekli Tehdit Tespiti, IEC 62443, Türk Dünyası, Siber Olgunluk Modeli, MITRE ATT&CK

I. GİRİŞ

Kritik enerji altyapıları; SCADA (Supervisory Control and Data Acquisition) ve ICS (Industrial Control System) bileşenlerinin entegrasyonu yoluyla izlenmekte ve denetlenmektedir. Bu sistemlerin işleyişinin bozulması; ulusal güvenliği, kamu güvenliğini ve ekonomik istikrarı doğrudan tehdit eden sonuçlar doğurmaktadır [1]. Türkiye özelinde yürütülen birinci el alan araştırmaları, ulusal kritik altyapıların yapısal düzeyde belirgin zafiyetler taşıdığını ortaya koymuştur: kamu ihale mevzuatının güvenlik tedarikine etkisi, kalifiye iş gücü açığı, üçüncü taraf hizmet sağlayıcılara ilişkin yetkilendirme süreçlerindeki belirsizlikler ve sektörel teknik kılavuz dokümanlarının eksikliği bu zafiyetlerin başlıcaları arasındadır [1].

Operasyonel teknoloji (OT) ile bilgi teknolojisi (IT) ağlarının yakınsaması saldırı yüzeyini genişletmektedir [2]. SCADA ortamlarında yaygın biçimde kullanılan haberleşme protokolleri; tasarımları gereği zayıf veya hiç bulunmayan kimlik doğrulama ve şifreleme mekanizmaları nedeniyle dış ağ kaynaklı tehditlere karşı savunmasız kalmaktadır [3]. Modbus protokolü erişim denetim mekanizması içermemekte ve haberleşme süreci şifrelenmemiş biçimde gerçekleşmektedir; bu durum, ağa erişimi olan herhangi bir aktörün meşru cihaz davranışı sergileyerek veri yakalama veya enjeksiyon saldırısı gerçekleştirmesine olanak vermektedir [3]. DNP3 protokolüne ait güvenlik mekanizmaları sınırlı düzeydedir; OPC UA üzerinden sağlanan IIoT iletişimi ise varsayılan yapılandırmalarda kimlik bilgilerinin açık metin halinde aktarılmasına yol açabilmektedir [4].

Enerji sektörüne yönelik kayıtlı olayların analizi, saldırı vektörlerinin büyük ölçüde kimlik avı (spear phishing), dışarıdan gerçekleştirilen kötü amaçlı yazılım enjeksiyonu, ağa açık alan cihazlarına doğrudan müdahale ve tedarik zinciri üzerinden yürütülen saldırılar etrafında yoğunlaştığını göstermektedir [5]. Ağ segmentasyonu eksikliği, kullanım ömrünü doldurmuş ekipmanların yerinde kalmaya devam etmesi ve yerel SCADA hatlarında şifrelenmemiş Modbus/TCP iletişiminin sürdürülmesi; söz konusu tehditlerin etki düzeyini büyütmektedir [5].

Tehdit modelleme tarafında MITRE ATT&CK çerçevesi, hasım davranışlarının taktik, teknik ve prosedür (TTP) ekseninde sınıflandırılması için kapsamlı bir bilgi tabanı sunmaktadır [6]. 2020 yılında yayımlanan ATT&CK for ICS alanı; OT ortamlarındaki post-compromise davranışları 11 taktik ve 78 teknik ile karakterize etmektedir [2]. Inhibit Response Function ve Impair Process Control taktikleri; doğrudan ICS ortamlarına özgüdür ve IT alanından OT alanına geçişi hedefleyen hasım davranışlarını kapsamaktadır [6].

Mevcut çerçeveler tek başına ele alındığında, her biri yalnızca belirli bir değerlendirme boyutuna

odaklanmaktadır. NIST CSF; tanımla, koru, tespit et, müdahale et ve kurtar olmak üzere beş işlev altında örgütlenen yüksek seviyeli bir risk yönetim yapısı sunmakla birlikte, kuruluş içi uygulama olgunluğunu ölçmek üzere tasarlanmamıştır; çerçevenin Tier yapısı bir olgunluk göstergesi değil yönetimsel bir profilleme aracıdır [7]. IEC 62443 standardı; IACS bileşenleri için SL1'den SL4'e kadar dört seviyeli bir güvenlik kapasitesi ve zone/conduit kavramını sağlamakta, ancak değerlendirme uygulayıcı yorumuna geniş bir alan bırakmaktadır [8]. MITRE ATT&CK ise hasım davranışını modellemekte fakat olgunluk skorlamasına dönük yapılandırılmış bir araç içermemektedir [6]. Diğer taraftan makine öğrenmesi tabanlı anomali tespiti yöntemleri yüksek doğruluk oranlarına ulaşmaktadır; sembolik dinamik filtreleme (SDF) ile özellik çıkarımı uygulayan ve Derin İnanç Ağı (DBN) mimarisi üzerine kurulan bir denetimsiz öğrenme sistemi, IEEE 39, 118 ve 2848 baralı test sistemlerinde %99 doğruluk ve %2'nin altında yanlış pozitif oranı raporlamaktadır [9]. Bununla birlikte, söz konusu yöntemlerin standart bir olgunluk çerçevesi içine yerleştirilmesine yönelik tümleşik bir öneri literatürde sınırlıdır.

Bu çalışmada; Türk Devletleri Teşkilatı bünyesindeki kritik enerji altyapılarının siber güvenlik olgunluğunu karşılaştırılabilir biçimde değerlendirebilmek üzere, IEC 62443 SL1–SL4 katmanlarını NIST CSF işlev ve kategori yapısı ile eşleştiren, MITRE ATT&CK for ICS taktik-teknik kaplama oranını ölçüt olarak kullanan ve denetimsiz öğrenmeye dayalı bir skorlama bileşeni içeren hibrit bir karşılaştırmalı çerçeve önerilmektedir. Türkiye için [1]'de raporlanan birinci el bulgular üzerinden çerçevenin uygulanış biçimi gösterilmiş, yöntemin Kazakistan, Azerbaycan, Türkmenistan ve Özbekistan'a uygulanması için prosedürel bir iş akışı tanımlanmıştır.

Bildirinin geri kalanı şu biçimde düzenlenmiştir: II. bölüm önerilen çerçevenin yöntemsel bileşenlerini tanımlamaktadır. III. bölüm Türkiye gösterimi üzerinden elde edilen bulguları sunmaktadır. IV. bölüm sonuçları, sınırlılıkları ve gelecekteki çalışma yönelimlerini ele almaktadır.

II. ÖNERİLEN ÇERÇEVE VE METODOLOJİ

A. Çerçeve Mimarisi ve Veri Kaynakları

Önerilen çerçeve; (i) tehdit yüzeyi haritalaması, (ii) standart tabanlı olgunluk skorlaması ve (iii) yapay zekâ destekli anomali algılama katmanı olmak üzere üç düzeyli bir mimariden oluşmaktadır. Mimarinin amacı; açık erişimli bilgi kaynaklarını — MITRE ATT&CK for ICS bilgi tabanı, CISA tarafından yayımlanan ICS-CERT uyarıları ve Ulusal Zafiyet Veritabanı (NVD) — tek bir olgunluk değerlendirmesi içinde tutarlı biçimde birleştirmektir.

Çerçevenin metodolojik temeli, Karabacak ve diğerlerinin (2016) Türkiye'nin kritik altyapı korumasında zafiyetin kök nedenlerini çıkarmak için kullandığı iki aşamalı yaklaşıma dayanmaktadır: birinci aşamada temellendirilmiş kuram (grounded theory) ile ulusal düzey siber güvenlik projelerinden zafiyet kök nedenlerinin çıkarılması, ikinci aşamada Delphi anketi yoluyla olgunluk kriterlerinin türetilmesi [1]. Bu çalışmada anılan şablon korunmakla birlikte; tek ülkeli bir değerlendirme yerine, beş ülkeye karşılaştırılabilir biçimde uygulanabilen bir ölçüm matrisi türetilmesi hedeflenmiştir.

B. Sistem Tanımı ve Zone/Conduit Modeli

İlk adımda, değerlendirilecek sistemin (System under Consideration — SuC) sınırları IEC 62443-3-2 yönergesinde tanımlandığı biçimde belirlenmektedir [4]. Enerji altyapısının operasyonel ortamı; saha cihazları (Düzey 0–1), temel denetim (Düzey 1), denetleyici denetim/SCADA (Düzey 2), operasyon yönetimi (Düzey 3) ve kurumsal sistemler (Düzey 4) olmak üzere PERA (Purdue Enterprise Reference Architecture) tabanlı beş düzeyli referans modele yerleştirilmektedir [8]. Her düzey; kritiklik seviyesine göre güvenlik bölgelerine (zone) ayrılmakta; bölgeler arası bilgi akışı conduit'ler aracılığıyla tanımlanmaktadır [4]. Modbus/TCP, DNP3 ve OPC UA protokolleri üzerinden gerçekleşen iletişim hatları bu yapıda birer conduit olarak ele alınmakta ve her conduit için hedef güvenlik seviyesi SL(target) ∈ {SL1, SL2, SL3, SL4} olarak atanmaktadır. Çerçevenin temel hedefi her bölge için SL(achieved) ≥ SL(target) koşulunun sağlanmasıdır [8].

C. Protokol–Tehdit Eşlemesi ve MITRE ATT&CK for ICS Sınıflandırması

İkinci adımda, SCADA protokollerinde belgelenmiş yapısal zafiyetler kayıt altına alınmaktadır. Modbus protokolünün erişim denetim mekanizması içermemesi ve haberleşmenin şifrelenmemiş gerçekleşmesi; DNP3 protokolünde tanımlanmış güvenlik mekanizmalarının sınırlı kalması; OPC UA için varsayılan yapılandırmalarda kimlik bilgilerinin açık metin biçiminde aktarılması [3], [4] zafiyet kümesinin temel bileşenleridir. Söz konusu zafiyetler; MITRE ATT&CK for ICS matrisindeki tekniklerle bire bir eşleştirilmektedir. ATT&CK for ICS, 11 taktik ve 78 teknik içermekte; Inhibit Response Function ve Impair Process Control taktikleri ise doğrudan OT ortamlarına özgü davranışları kapsamaktadır [2], [6].

Eşleme sonucunda her bir ülke c için bir tehdit profili vektörü T_c oluşturulmaktadır:

$$T_c = \{ (p_i, v_j, \tau_k, SL_{target}) \mid p_i \in P, v_j \in V, \tau_k \in ATT\&CK_{ICS} \} \quad (1)$$

Burada P protokol kümesini, V belgelenmiş zafiyet kümesini, τ_k ise ilgili zafiyetin operasyonel gerçeklemesinde kullanılabilecek ATT&CK tekniğini ifade etmektedir.

D. NIST CSF × IEC 62443 SL Olgunluk Skorlama Matrisi

Üçüncü adımda, T_c üzerinden ülkenin olgunluğu iki eksenli bir matriste puanlanmaktadır. Satırlar NIST CSF'nin beş işlevini (Tanımla, Koru, Tespit Et, Müdahale Et, Kurtar) [7]; sütunlar IEC 62443'ün SL1–SL4 güvenlik seviyelerini [8] oluşturmaktadır. Almuhammadi ve Alsaleh tarafından önerilen ve NIST CSF kategorilerine uyum değerlendirmesini ekleyen beş düzeyli ölçek (0–4) bu çalışmada da hücre düzeyinde benimsenmiştir [7]. Her hücre için olgunluk skoru $m_{f,SL} \in [0, 4]$ aralığında atanmakta ve ülke düzeyi toplam olgunluk skoru aşağıdaki biçimde hesaplanmaktadır:

$$M_c = \Sigma_f \Sigma_{SL} w_{f,SL} \cdot m_{f,SL} \quad (2)$$

Ağırlık terimi w_{f,SL}; her conduit için belirlenmiş SL(target) ile gözlemlenen SL(achieved) arasındaki açık dikkate alınarak atanmakta; kritik SL seviyelerine ait hücrelere ağırlık öncelik tanınmaktadır.

E. Denetimsiz Öğrenmeye Dayalı Yapay Zekâ Bileşeni

Dördüncü adımda, anomali algılama yetkinliği bir olgunluk göstergesi olarak çerçeveye entegre edilmektedir. Karimipour ve diğerlerinin (2019) önerdiği denetimsiz öğrenme yaklaşımı; sembolik dinamik filtreleme (SDF) tabanlı özellik çıkarımı ve Derin İnanç Ağı (DBN) mimarisi üzerine kurulmuştur. Yöntem etiketsiz veri ile çalışmakta ve anomali göstergesi olarak serbest enerji değerini kullanmaktadır. IEEE 39, 118 ve 2848 baralı test sistemleri üzerinde %99 doğruluk, %98 gerçek pozitif oranı ve %2'nin altında yanlış pozitif oranı raporlanmıştır [9].

Söz konusu yaklaşımın bir ülke ölçeğinde uygulanabilirlik düzeyi üç ölçüt üzerinden değerlendirilmektedir: (i) ülkede konuşlu enerji kuruluşlarında ML tabanlı anomali algılama sisteminin operasyonel varlığı, (ii) raporlanan yanlış pozitif oranı, (iii) anomali tespitinin SCADA halkasına dahil edilme süresi (gerçek zamanlı/toplu işleme). Bu üç boyut; NIST CSF'nin "Tespit Et" işlevinin SL3 ve SL4 hücrelerine doğrudan katkı sağlamaktadır.

F. Bölgesel Olgunluk Endeksi

Beşinci ve son adımda, Türkiye, Kazakistan, Azerbaycan, Türkmenistan ve Özbekistan için M_c değerleri aynı w_{f,SL} ağırlık matrisi kullanılarak hesaplanmakta ve karşılaştırılabilir bir bölgesel olgunluk endeksi oluşturulmaktadır. Endeks; ülke düzeyinde tek bir sayısal değer üretmekle birlikte, alt bileşenlere geri dönüşle protokol bazlı, NIST CSF işlevi bazlı ve SL bazlı ayrıntılı analizlere de olanak sağlamaktadır. Açık erişimli kaynaklardan veri çekilemeyen hücreler "veri yok" olarak işaretlenmekte ve toplam skorun istatistiksel anlamlılığı raporlanan değerler üzerinden hesaplanmaktadır.

III. BULGULAR

A. Türkiye İçin Çerçevenin Veri Beslemesi ve Hücre Skorlaması

Türkiye'nin değerlendirilmesinde Karabacak ve diğerlerinin (2016) Delphi anketinden elde ettiği birinci el sayısal değerler doğrudan girdi olarak kullanılmıştır [1]. Söz konusu çalışmada uzman görüşüne dayalı olarak 40 olgunluk kriteri puanlanmış; bu kriterler önerilen çerçevenin NIST CSF işlev ve IEC 62443 SL eksenli hücrelerine eşlenmiştir.

NIST CSF "Tanımla" işlevinin alt hücreleri için elde edilen örnek girdiler Tablo I'de sunulmuştur. Tabloda; düzenleyici çerçeveye ilişkin yüksek-düzey kriterlerin (SL1) görece güçlü bir profil sergilediği (≥ 2.33), sektör-spesifik teknik kılavuzluk düzeyinde (SL3) ise zayıf bir konum oluştuğu (≤ 1.83) gözlemlenmektedir.

TABLO I.TÜRKİYE İÇİN "TANIMLA" İŞLEVİ ALT HÜCRELERİNİN ÖRNEK PUANLAMASI ([1]'DE RAPORLANAN DELPHİ VERİSİ)

Kriter	SL	Puan
Asgari güvenlik karşı tedbirlerini zorunlu kılan düzenlemeler	SL1	2.50
Bilgi sistemleri ve karşı tedbirlerin niteliklerini belirleyen düzenlemeler	SL2	2.33
Sektöre özgü teknik kılavuz dokümanlar	SL3	1.50
Ulusal ve sektörel standartlar	SL3–SL4	1.83

Aynı eşleme yöntemi; "Koru", "Tespit Et", "Müdahale Et" ve "Kurtar" işlevleri için [1]'de raporlanan kalan 36 kriter üzerinden tekrarlanabilmektedir. Mevcut Delphi verisinin tek başına matrisin bütün hücrelerini kaplamadığı; dolayısıyla Türkiye için yapısal bütünlüğün korunmakla birlikte bazı hücrelerin ek alan çalışmasını gerektirdiği tespit edilmiştir.

B. Protokol Düzeyi Tehdit Profili

Pliatsios ve diğerlerinin (2020) SCADA protokol incelemesi; Modbus protokolünün açık spesifikasyonu ve donanım yaygınlığı nedeniyle en yaygın kullanılan SCADA protokolü olduğunu raporlamıştır [3]. Aynı protokolün erişim denetimi içermemesi ve şifrelenmemiş haberleşmesi; ağa erişimi olan bir aktörün meşru cihaz davranışı sergileyerek Adversary-in-the-Middle, sahte komut enjeksiyonu ve DoS saldırıları gerçekleştirmesine zemin oluşturmaktadır [3]. DNP3 üzerinde belgelenmiş açıklara karşı önerilmiş güvenlik çözümlerinin önemli bir kısmı; kural tabanlı veya anomali tespiti odaklı saldırı algılama yaklaşımları ile AES, RSA ve SHA-2 algoritmalarına dayalı şifreleme çözümleridir [3]. OPC UA protokolüne ait varsayılan IIoT iletişimi; kimlik bilgilerini açık metin biçiminde aktarabildiğinden sniffing saldırılarına açıktır ve kimlik doğrulama, şifreleme algoritması veya dijital imza gibi ek katmanlarla güçlendirilmediğinde yetkisiz erişim için kritik bir vektör oluşturmaktadır [4].

Bu zafiyetler MITRE ATT&CK for ICS matrisinde Initial Access, Lateral Movement, Inhibit Response Function ve Impair Process Control taktiklerine doğrudan eşlenmektedir [2], [6]. Söz konusu eşleme; çerçevenin (1) numaralı eşitliği aracılığıyla her ülke için tekrarlanabilir niteliktedir.

C. MITRE ATT&CK for ICS Kapsama Analizi

Al-Sada ve diğerlerinin (2024) MITRE ATT&CK veri tabanı üzerinde yürüttüğü istatistiksel analiz; ICS alanının 78 teknikten oluştuğunu ve yazılım profilleri tarafından nadiren hedeflendiğini ortaya koymaktadır [6]. Aynı çalışmada hasım gruplarının %43'ünün casusluk (espionage), %29'unun finansal motivasyonla hareket ettiği raporlanmıştır; finansal motivasyona sahip grupların ICS ve mobil platformları neredeyse hiç hedeflemediği ayrıca tespit edilmiştir [6]. Siber savaş/yıkım motivasyonu yalnızca beş hasım grubuna; politik veya ideolojik motivasyon ise dört hasım grubuna atfedilmiştir [6].

Bu bulgular; enerji altyapısı tehdit profilinin ağırlıklı olarak devlet destekli veya ideolojik motivasyonlu aktörler tarafından şekillendirildiğini, dolayısıyla bölgesel olgunluk değerlendirmesinde "Tespit Et" ve "Müdahale Et" işlevlerinin Advanced Persistent Threat (APT) odaklı kapasitelere ağırlık vermesi gerektiğini göstermektedir.

D. AI Bileşeninin Beklenen Performans Tabanı

Çerçevenin AI katmanına entegre edilen SDF + DBN modelinin referans performansı, Karimipour ve diğerlerinin (2019) IEEE 39, 118 ve 2848 baralı test sistemleri üzerinde gerçekleştirdiği deneylerden alınmaktadır: doğruluk %99, gerçek pozitif oranı %98 ve yanlış pozitif oranı %2'nin altı [9]. Bu değerler; söz konusu yöntemin "Tespit Et" işlevi altında SL3 ve SL4 hedeflerine ulaşmak için yeterli bir performans tabanı sağladığını ortaya koymaktadır.

Bununla birlikte söz konusu performansın ülke ölçeğinde gerçeklenebilmesi üç ön koşula bağlıdır: (i) telemetri verisinin yeterli sıklıkta toplanabilmesi, (ii) RTU ve PMU ağının homojen biçimde donatılmış olması ve (iii) makine öğrenmesi modelinin yerel yönetişim altında işletilebilmesi. Bu üç koşul, çerçevenin (2) numaralı eşitliğindeki "Tespit Et" satırına ait ağırlık terimlerinde doğrudan kalibrasyon bileşeni olarak kullanılmaktadır.

E. Türk Devletleri Teşkilatı'nın Diğer Üye Ülkeleri İçin Uygulanma Yol Haritası

Kazakistan, Azerbaycan, Türkmenistan ve Özbekistan'a ait birinci el ülke düzeyi telemetri veya olgunluk verisi; açık erişimli kaynaklarda Türkiye verisi ile eşit hacim ve ayrıntıda bulunmamaktadır. Bu durum; söz konusu ülkeler için çerçevenin uygulanamayacağı anlamına gelmemekte, aksine aşamalı bir veri toplama prosedürünün gerekliliğini ortaya koymaktadır. Belirlenen uygulanma adımları şu sırayı izlemektedir:

1) Açık kaynak veri çekimi: İlgili ülkelerde faaliyet gösteren enerji kuruluşlarının kamuya açık raporları, MITRE ATT&CK for ICS olay kayıtları ve CISA ICS-CERT uyarıları üzerinden başlangıç tehdit profilinin oluşturulması.
2) Delphi panelinin yerelleştirilmesi: Karabacak ve diğerlerinin (2016) Türkiye için uyguladığı yarı yapılandırılmış görüşme + Delphi anketi formatının söz konusu ülkelerin kritik altyapı paydaşlarına genişletilmesi [1].
3) Protokol envanteri: Ülke düzeyinde SCADA protokol dağılımının (Modbus/TCP, DNP3 ve OPC UA payları) belgelenmesi ve her conduit için SL(target) atanması [4], [8].
4) Hücre puanlaması: Toplanan girdilerin (1) ve (2) numaralı eşitlikler aracılığıyla her ülke için bir M_c değerine dönüştürülmesi.
5) Karşılaştırmalı endeks raporlaması: M_c değerlerinin tek bir bölgesel olgunluk tablosunda sunulması ve veri eksiği bulunan hücrelerin açıkça işaretlenmesi.

Bu yol haritası; çerçevenin gelecek araştırmalar için yeniden kullanılabilirliğini güvence altına almakta ve Türk Devletleri Teşkilatı düzeyinde ortak bir veri paylaşım protokolünün gerekliliğini açıkça ortaya koymaktadır.

IV. SONUÇ

Bu çalışmada; Türk Devletleri Teşkilatı bünyesindeki kritik enerji altyapılarının siber güvenlik olgunluğunu karşılaştırılabilir biçimde değerlendirmek üzere; IEC 62443 SL1–SL4 katmanlarını NIST CSF beş işlev yapısı ile eşleştiren, MITRE ATT&CK for ICS taktik-teknik kaplama oranını ölçüt olarak kullanan ve denetimsiz öğrenmeye dayalı bir skorlama bileşeni içeren hibrit bir karşılaştırmalı çerçeve önerilmiştir.

Çerçeve; Türkiye için Karabacak ve diğerlerinin (2016) Delphi anketinden elde edilen birinci el sayısal değerler [1] kullanılarak gösterimsel olarak uygulanmıştır. Elde edilen sonuçlar; Türkiye'nin üst-düzey düzenleyici çerçevede (SL1) orta düzeyde bir olgunluk profili sergilediğini, sektör-spesifik teknik kılavuzluk düzeyinde (SL3) ise belirgin bir zayıflık taşıdığını ortaya koymaktadır. SCADA protokollerinde belgelenmiş yapısal zafiyetler — Modbus/TCP'de erişim denetimi yokluğu, DNP3'te sınırlı güvenlik kontrolleri ve OPC UA'da açık metin kimlik aktarımı — MITRE ATT&CK for ICS matrisinin Initial Access, Lateral Movement, Inhibit Response Function ve Impair Process Control taktiklerine eşlenmiştir. AI bileşeni için Karimipour ve diğerlerinin (2019) raporladığı %99 doğruluk ve %2'nin altında yanlış pozitif oranı [9]; çerçevenin "Tespit Et" işlevi altında SL3 ve SL4 hedeflerine ulaşılabilirlik için yeterli bir performans tabanı oluşturmaktadır.

Çalışmanın temel sınırlılığı; Kazakistan, Azerbaycan, Türkmenistan ve Özbekistan için açık erişimli kaynaklarda Türkiye verisi ile eşit hacim ve ayrıntıda birinci el verinin bulunamamasıdır. Bu nedenle önerilen çerçeve; söz konusu dört ülke için ampirik olarak işletilmemiş, doldurulabilir bir prosedürel çerçeve olarak konumlandırılmıştır. AI bileşenine ait performans değerleri test sistemleri (IEEE 39, 118 ve 2848 baralı) üzerinde elde edilmiş olup, gerçek operasyonel ortamlarda kalibrasyon gerektirmektedir [9]. NIST CSF kategorilerinin yerel düzeyde adlandırılması ve IEC 62443 SL hedeflerinin sektörel düzeyde belirlenmesi de uygulayıcı yorumuna açık alanlar olarak kalmaktadır.

Önerilen çerçevenin tam ampirik biçimde işletilebilmesi için Türk Devletleri Teşkilatı düzeyinde paylaşımlı bir veri toplama protokolüne ihtiyaç vardır. Söz konusu protokol; (i) yerelleştirilmiş Delphi panellerinin eş zamanlı uygulanmasını, (ii) SCADA protokol envanterlerinin standartlaştırılmış biçimde derlenmesini ve (iii) AI tabanlı anomali tespiti sistemlerinin operasyonel performans verilerinin merkezi biçimde raporlanmasını kapsamalıdır. Çerçevenin denetimli öğrenme yöntemleriyle (örneğin SVM ve Karar Ağaçları temelli sınıflandırıcılar [3]) genişletilmesi ve organizasyonel ile bireysel düzeyde siber güvenlik kültürü boyutunun [10] çerçeveye eklenmesi sonraki çalışma yönelimleri arasındadır.

Kaynakça

[1] B. Karabacak, S. Ozkan Yildirim, and N. Baykal, "A vulnerability-driven cyber security maturity model for measuring national critical infrastructure protection preparedness," International Journal of Critical Infrastructure Protection, vol. 15, pp. 47–59, 2016.

[2] B. Al-Sada, A. Sadighian, and G. Oligeri, "MITRE ATT&CK: State of the Art and Way Forward," ACM Computing Surveys, vol. 57, no. 1, Article 12, pp. 1–37, Oct. 2024.

[3] D. Pliatsios, P. Sarigiannidis, T. Lagkas, and A. G. Sarigiannidis, "A Survey on SCADA Systems: Secure Protocols, Incidents, Threats and Tactics," IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1942–1976, 2020.

[4] H. Lalaoui Hassani, A. Bahnasse, E. Martin, C. Roland, O. Bouattane, and M. El Mehdi Diouri, "Vulnerability and security risk assessment in a IIoT environment in compliance with standard IEC 62443," Procedia Computer Science, vol. 191, pp. 33–40, 2021.

[5] G. Stergiopoulos, D. A. Gritzalis, and E. Limnaios, "Cyber-Attacks on the Oil & Gas Sector: A Survey on Incident Assessment and Attack Patterns," IEEE Access, vol. 8, pp. 128440–128475, 2020.

[6] B. Al-Sada, A. Sadighian, and G. Oligeri, "Analysis and Characterization of Cyber Threats Leveraging the MITRE ATT&CK Database," IEEE Access, vol. 12, pp. 1216–1234, 2024.

[7] S. Almuhammadi and M. Alsaleh, "Information Security Maturity Model for NIST Cyber Security Framework," in Computer Science & Information Technology (CS & IT), 2017, pp. 51–62.

[8] B. Leander, A. Čaušević, and H. Hansson, "Applicability of the IEC 62443 standard in Industry 4.0 / IIoT," in Proc. 14th International Conference on Availability, Reliability and Security (ARES '19), Canterbury, U.K., Aug. 2019, pp. 1–8.

[9] H. Karimipour, A. Dehghantanha, R. M. Parizi, K.-K. R. Choo, and H. Leung, "A Deep and Scalable Unsupervised Machine Learning System for Cyber-Attack Detection in Large-Scale Smart Grids," IEEE Access, vol. 7, pp. 80778–80788, 2019.

[10] A. Georgiadou, S. Mouzakitis, and D. Askounis, "Assessing MITRE ATT&CK Risk Using a Cyber-Security Culture Framework," Sensors, vol. 21, no. 9, Article 3267, May 2021.

Akıllı Şehir Tedarik Zincirlerinde Yapay Zeka Tabanlı Yazılım Bileşen Analizi: Türk Dünyası Belediyeleri İçin Otonom Zafiyet Tarama Yaklaşımı

Taha Tokgöz
Yönetim Bilişim Sistemleri Bölümü
Ankara Medipol Üniversitesi
Ankara, Türkiye
taha@tahatokgoz.com

Cevat Özarpa
Siber Güvenlik Anabilim Dalı
Ankara Medipol Üniversitesi
Ankara, Türkiye
cevat.ozarpa@ankaramedipol.edu.tr

Özet—Bu araştırmada, Türk Dünyası'ndaki şehirlerin akıllı dönüşüm süreçlerinde kullandıkları yazılımların tedarik zinciri güvenliğine ışık tutuluyor; aynı zamanda SBOM'a dayalı otomatik bir zafiyet denetimi yöntemi öne sürülüyor. İncelenen altyapılarda yer alan IoT cihazlarının firmware'leri, açık kaynak kodlu kütüphaneler ve buluta yayılmış hizmet yapıları üzerinde duruluyor - örneklem olarak İstanbul, Bişkek, Bakü ile Astana seçilmiştir. Güvenlik açıkları, OWASP SCVS modeli yardımıyla sınıflandırılıyor; bu sayede yazılım parçalarının ne kadar sızdıraç barındırdığı haritalanıyor. Mevcut bilimsel üretim genellikle büyük ölçekli şirket merkezli sistemlere yöneliyor, oysa belediye temelli dijitalleşme girişimlerindeki karışık yapılı bileşenler için kapsamlı incelemeler nadiren karşılaşılmakta, özellikle kültürel coğrafya bağlamında böyle bir boşluk belirginleşiyor. Çalışma süresince dört kentin akıllı yönetim platformlarında hangi özgür yazılımların işlediği ayırt edildi; kamu açıklık verilerinden yararlanarak mevcut güvenlik riskleri listelendi, ayrıca önerilen SBOM destekli tarayıcı mantığının teknik gerçeklenebilirliği tartışıldı. Bazı akıllı şehir girişimlerinde sıklıkla görülen açık kodlu parçalarda yer alan güvenlik açıklarının yoğunluğu gösterilmiştir. Bir yandan SBOM'a dayalı otonom denetleme yönteminin belediye bağlamında ne kadar işe yarayabileceği ortaya çıkarılmıştır. Diğer yanda bu yöntemle karşılaşılacak teknik engeller haritalanmıştır. Araştırma, yazılım sağlayıcı güvenliği üzerine yapılan ilk kapsamlı karşılaştırma niteliğindedir. Özellikle Türk Dünyası'ndaki yerel yönetimler için geçerlidir durum. Akıllı kent projelerinde otomatik açlık tarayıcısının pratik kullanımı incelenmiştir.

Anahtar Kelimeler— Akıllı Şehir Güvenliği, Yazılım Tedarik Zinciri, SBOM, Otonom Zafiyet Tarama, Türk Dünyası, IoT Güvenliği, Açık Kaynak Bileşen Analizi

I. GİRİŞ

Akıllı şehir kavramı, kentsel hizmetlerin sensörler, ağ altyapısı, bulut tabanlı veri işleme ve uygulama katmanları üzerinden bütünleştirilmesini ifade eder [1], [2]. ITU-T Y.4000/Y.2060 referans modeli akıllı şehir IoT mimarisini cihaz, ağ, hizmet ve uygulama olmak üzere dört katmana ayırır [1]. Bu mimari; akıllı yönetim, akıllı ulaşım, akıllı çevre, akıllı sağlık, akıllı yaşam, akıllı ekonomi ve akıllı insan olarak adlandırılan domenlere hizmet vermektedir [3]. Belediye düzeyindeki dijital dönüşüm süreçleri, söz konusu katmanları oluşturan donanım ve yazılım bileşenlerini farklı tedarikçilerden temin etmek zorundadır; bu heterojenlik, tekil bir güvenlik politikasının uygulanmasını ve tutarlı modellemeyi engeller [1], [4]. Snap4City gibi referans platformlar dahi IoT Edge cihazlarda Raspberry Pi, Linux, Windows, Docker veya Android tabanlı çoklu çalışma zamanlarını barındırmakta; bu yapı saldırı yüzeyini doğrudan genişletmektedir [4].

Yazılım tedarik zinciri, modern uygulama gereksinimlerini karşılamak üzere bir araya getirilen ikili dosyalar, kütüphaneler, araçlar ve mikroservislerden oluşan büyüyen bir bileşenler kümesidir [5]. Üretici ürünleri açık kaynak ve ticari bileşenlerin birleşiminden meydana gelmekte; tek bir bileşendeki açıklık birden çok ürüne yayılma potansiyeli taşımaktadır [5]. Sonatype 2021 raporu yazılım tedarik zinciri saldırılarında yıllık %650 artış kaydederken [5], 2019–2022 döneminde bu oran %742 olarak ölçülmüştür [6]. SolarWinds saldırısında Orion yazılımı trojanize edilmiş bir güncelleme aracılığıyla çok sayıda kamu kurumunu etkilemiş; redis-py açık kaynak kütüphanesindeki bir kusur 2023'te ChatGPT kullanıcı sohbet geçmişlerinin sızmasına neden olmuştur [6]. Akıllı şehir bağlamında somut bir saldırı örneği olarak, SCADA tabanlı akıllı trafik lambası kontrol sistemlerine yönelik Malware, Buffer Overflow, DoS ve Jamming saldırıları belgelenmiş; bu sistemlerin güvenliği için Ağ-Tabanlı ve Ana-Bilgisayar-Tabanlı Saldırı Tespit Sistemlerinin (IDS) entegrasyonu önerilmiştir [13]. Bu olaylar, kritik altyapı statüsündeki belediye sistemlerinde benzer senaryoların yıkıcı sonuçlar doğurabileceğini göstermektedir.

IoT cihaz firmware'leri üçüncü taraf bileşenlere (Third-Party Components, TPC) yoğun bağımlılık göstermektedir. Zhao ve diğerleri tarafından 34.136 firmware imajı üzerinde gerçekleştirilen büyük ölçekli ampirik analiz, 584 TPC tespit etmiş ve 429 CVE'ye karşılık gelen toplam 128.757 güvenlik açığı raporlamıştır [7]. Aynı çalışma, firmware'lerde kullanılan TPC sürümlerinin güncel hâllerinin ortalama beş yıl gerisinde kaldığını ve cihazların güvenlik durumunun coğrafi dağılıma göre farklılaştığını ortaya koymaktadır [7]. Belediye akıllı şehir altyapılarında ise OWASP Top 10 kapsamındaki SQL Injection, Broken Authentication, Sensitive Data Exposure, Broken Access Control, Security Misconfiguration ve Using Components with Known Vulnerabilities sınıfları doğrudan görülmekte; özellikle son sınıf, tedarik zinciri açıklıkları ile birebir örtüşmektedir [4].

Belediye akıllı şehir sistemlerinin uluslararası güvenlik standartlarına uyumu, kritik altyapı güvenliğinin yapı taşıdır. 1971-2020 döneminde geliştirilen 31 uluslararası güvenlik standardının taraması, bu standartların %71'inin (22 adet) ABD kaynaklı olduğunu ve standartlara uyum derecesinin sistem güvenlik seviyesiyle doğru orantılı seyrettiğini ortaya koymaktadır

[12]. Kritik enerji altyapıları özelinde 17 risk türü üzerinden yürütülen değerlendirme, siber saldırı riskini güvenlik riski (%8,411) ve personel riski (%7,019) sıralamasında üçüncü konuma (%8,005) yerleştirmektedir [14]. Bu çerçevede yenilenebilir enerji kaynakları %9 risk taşırken elektrik ve doğalgaz altyapıları sırasıyla %13 ve %15 risk seviyesi göstermektedir [14]. Söz konusu bulgular, belediye akıllı şehir bileşenlerinin yazılım tedarik zinciri güvenliğinin standartlar zemininde değerlendirilmesi gerektiğini desteklemektedir.

Yazılım Malzeme Listesi (Software Bill of Materials, SBOM), bir uygulamayı oluşturan tüm bileşenlerin ve bağımlılık ilişkilerinin makineler tarafından okunabilir envanterini sunan ve hızla yazılım tedarik zinciri güvenliğinin köşe taşına dönüşen bir teknolojidir [5], [6]. Amerika Birleşik Devletleri'nin 14028 sayılı İcra Emri, federal hükümete yazılım sağlayan tedarikçilere SBOM teslim yükümlülüğü getirmekte ve NTIA standartlarına uyumu zorunlu kılmaktadır [5]. SBOM'un statik analiz araçlarıyla birleştirilmesi tedarik zincirindeki açıklıkların tespiti için uygulanabilir bir yaklaşım sunmakta [5]; Trivy ve Grype kapsayıcı imajlarındaki paketleri taramakta, OWASP Dependency-Check ve Eclipse Steady ise Java/JavaScript ekosistemleri için yazılım kompozisyonu analizi (Software Composition Analysis, SCA) yürütmektedir [5], [8]. Cruz ve diğerleri, açık kaynak güvenlik değerlendirme araçlarını SAST, SCA ve DAST kategorileri altında karşılaştırmış; OWASP Top 10 kapsama, CI/CD entegrasyonu, SARIF birlikte çalışabilirliği ve çerçeve farkındalığı boyutlarında bir tipoloji üretmiştir [9]. Buna karşın SBOM paylaşımı; verilerin değiştirilebilirliği, satıcıların tam bilgi açıklamadaki çekinceleri ve paydaşlar arası uyumsuz gereksinimler nedeniyle sınırlanmaktadır [6].

Yapay zekâ tabanlı yöntemler, firmware ve yazılım bileşeni analizinde giderek artan bir rol üstlenmektedir. Derin öğrenme yapıları yazılım açıklıklarının tespiti için sinir ağı temelli yaklaşımlar sağlamakta; federe öğrenme (FL) ise heterojen mimariler üzerinde dağıtık model eğitimine olanak tanımaktadır [10]. Statik analizde edit distance ve kontrol akış grafiği (Control-Flow Graph, CFG) eşleme tekniklerinin struct2vec gibi gömülü ağ tabanlı benzerlik modelleriyle birleştirilmesi, TPC'lerin sürüm seviyesinde tanımlanma doğruluğunu yükseltmektedir [7]. Bu altyapı, imza tabanlı geleneksel taramaya kıyasla otonom denetim yetenekleri sağlamakta; ML-destekli emülasyon ve blok zinciri destekli federe öğrenme, küresel zafiyet sınıflandırma modellerinin sürekli güncellenmesi için somut bir çerçeve önermektedir [10]. Yapay zekâ uygulamalarının akıllı şehir siber güvenliği bağlamında entegrasyonu, otonom denetim yetenekleri için temel oluşturmaktadır [12], [14].

Mevcut literatürün belirgin sınırlılığı, yazılım tedarik zinciri ve TPC analizine yönelik çalışmaların büyük çoğunluğunun kurumsal yazılım geliştirme ekosistemlerine (npm, Maven, Docker depoları, açık kaynak Java projeleri) odaklanmış olmasıdır [5], [8]. Belediye düzeyinde akıllı şehir dağıtımlarında çalışan heterojen yazılım yığınlarının bütüncül incelemesi sınırlı kalmakta; politikacılar ve şehir yöneticilerinin IoT bileşenlerinin güvenlik olgunluğunu ölçebileceği metodolojik bir çerçeve eksikliği akademik literatürde açıkça raporlanmaktadır [1]. Türk Dünyası'ndaki yerel yönetimler — İstanbul, Bakü, Astana ve Bişkek gibi metropoller — akıllı şehir dönüşümlerine kaynak ayırmasına rağmen, bu kentlerin yazılım tedarik zinciri güvenliğine yönelik kavramsal düzeyde ortak bir karşılaştırmalı değerlendirme çerçevesi mevcut literatürde bulunmamaktadır. Yönetsel, dilsel ve düzenleyici açılardan benzerlikler taşıyan bu belediyeler için, ITU-T katman modeli ile OWASP SCVS sınıflandırmasını SBOM tabanlı otonom tarama mantığına entegre eden ortak bir metodolojiye ihtiyaç vardır.

Bu çalışmanın katkıları aşağıdaki gibidir:

1) ITU-T IoT katman modeli, OWASP SCVS sınıflandırması, SBOM teknolojisi ve AI/ML destekli benzerlik analizi unsurlarını entegre eden, belediye ölçeğine uyarlanmış kavramsal bir otonom zafiyet tarama çerçevesinin önerilmesi;
2) Çerçevenin Türk Dünyası'nda dört temsilî metropol (İstanbul, Bakü, Astana, Bişkek) bağlamında uygulanabilirlik koşullarının literatür temelli senaryo analizi ile değerlendirilmesi;
3) SBOM tabanlı denetimin belediye seviyesindeki teknik engellerinin — heterojen mimari ve birden çok firmware yığını desteği, tedarikçi şeffaflık eksikliği, SBOM'ların değiştirilebilirliği, AI/ML modellerinin coğrafi/dilsel veri çeşitliliği gereksinimi — sistematik olarak haritalandırılması; ayrıca çerçevenin Olgunluk Düzeyi değerlendirmesi ile genişletilmesi için AHP tabanlı kavramsal hat sunulması [15].

Bildirinin geri kalanı şu şekilde yapılandırılmıştır: Bölüm II önerilen metodolojiyi sunar. Bölüm III literatür temelli senaryo analiziyle elde edilen bulguları sergiler. Bölüm IV sonuç ve gelecek çalışma yönelimlerini tartışır.

II. ÖNERİLEN METODOLOJİ

A. Araştırma Tasarımı

Çalışma, akıllı şehir tedarik zinciri güvenliğini belediye ölçeğinde değerlendirebilecek bir otonom denetim çerçevesinin tasarım bilimi (design science) yönelimiyle geliştirilmesini ve literatür temelli senaryo analizi ile değerlendirilmesini kapsamaktadır. Çerçeve, ITU-T Y.4000/Y.2060 IoT katman modeli üzerine kurgulanmış [1] ve dört aşamalı bir işleme hattı olarak tanımlanmıştır: (i) varlık ve hizmet envanterinin çıkarılması, (ii) SBOM üretimi, (iii) AI/ML-destekli bileşen ve zafiyet eşlemesi, (iv) risk sınıflandırma ve raporlama. Çerçevenin değerlendirilmesi, Türk Dünyası'nda akıllı şehir dönüşümü yürüten dört temsilî metropol — İstanbul, Bakü, Astana ve Bişkek — kullanılarak senaryo bazında yapılmıştır. Belediyelere ait birincil yazılım envanteri verisi mevcut bilgi tabanında bulunmadığından, şehir-spesifik ampirik bulgu üretilmemiş; bunun yerine kontrollü senaryo değişkenleri üzerinden uygulanabilirlik koşulları haritalandırılmıştır. Çerçeveye eklenecek olgunluk değerlendirmesi için Analytic Hierarchy Process (AHP) yöntemi, çoklu kriter karar verme aracı olarak metodolojik altlık sağlamaktadır; Özarpa ve diğerlerinin Türkiye'de 37 siber güvenlik uzmanı ile yürüttüğü AHP destekli olgunluk çalışması bu çalışmada yöntemsel referans olarak alınmıştır [15].

B. Önerilen Çerçevenin Mimari Tanımı

Faz 1 — Varlık ve Hizmet Envanteri. ITU-T modeline göre belediye akıllı şehir sisteminin cihaz, ağ, hizmet ve uygulama katmanlarındaki tüm bileşenler enumerate edilir [1]. Hizmet katmanında akıllı yönetim, akıllı ulaşım, akıllı çevre, akıllı sağlık, akıllı yaşam, akıllı ekonomi ve akıllı insan domenleri ayrı ayrı işaretlenir [3]. Cihaz katmanında IoT Edge cihazları — Raspberry Pi, Linux/Windows tabanlı bilgisayarlar, Docker konteynerleri, Android tabanlı el cihazları gibi — heterojen çalışma zamanlarına göre etiketlenir [4].

Faz 2 — SBOM Üretimi. Her varlık için CycloneDX veya SPDX formatında SBOM üretilir [6]. Konteynerleştirilmiş hizmetler için Trivy ve Grype'ın paket çıkarım kabiliyeti kullanılır [5], [9]. Kaynak kod düzeyinde çözümleme yapılacak hizmetler için DevSecOps döngüsünün pre-build, build ve binary aşamalarına denk gelen üç tarama modu eşzamanlı çalıştırılır [8]. IoT firmware imajları, FirmSec'te tanımlanan ön işleme zinciri ile (paket yapısı çözme, sözdizimsel özellik çıkarımı, montaj çözme) bileşen sınırlarına ayrıştırılır [7]. SBOM'lar bağımlılık derinliği, doğrudan/transitif bağımlılık ayrımı, paket sürüm aralığı ve geliştirici/tedarikçi meta verisi içerecek şekilde zenginleştirilir [11].

Faz 3 — AI/ML-Destekli Bileşen ve Zafiyet Eşlemesi. Bileşenlerin tanımlanması iki katmanlı bir benzerlik analizine dayanır. Sözdizimsel eşleme katmanında, üçüncü taraf bileşenden çıkarılan özellikler ile hedef yazılım/firmware'den çıkarılan özellikler arasında edit distance metriği hesaplanır; ratio tabanlı eşleme oranı kullanılır [7]. FirmSec değerlendirmesinde elde edilen 0,74 ve 0,52 eşik değerleri başlangıç parametreleri olarak alınır ve belediye bağlamında ileri kalibrasyona açık tutulur [7]. Anlamsal eşleme katmanında, çıkarılan kontrol akış grafikleri (CFG) struct2vec gömülü sinir ağı ile vektörleştirilir; saldırıya açık fonksiyon vektörleri ile firmware fonksiyon vektörleri arasındaki kosinüs benzerliği hesaplanır ve en benzer üst-K fonksiyon adayları belirlenir [7]. Bileşen tespiti sonrasında her bileşenin sürüm-seviyesi tanımlaması, sadece o sürüme özgü sözdizimsel imzalar üzerinden gerçekleştirilir [7]. AI/ML bileşeninin genişletilmesi için derin öğrenme ve federe öğrenme mimarileri öngörülmekte; coğrafi olarak dağıtık belediyelerin yerel modellerinin merkezi olmayan biçimde birleştirilmesi AIBOM uzantısı ile mümkün kılınmaktadır [6], [10].

Faz 4 — Risk Sınıflandırma ve Raporlama. Eşlenen her bileşen için CVE ve GitHub Security Advisories (GHSA) veri tabanları ile çapraz sorgulama yapılır [5]. Zafiyet ciddiyeti, Common Vulnerability Scoring System (CVSS) v3 ölçeği — Low, Moderate, High, Critical — temelinde derecelendirilir [5], [11]. Sömürü kanıtı durumu (public exploit, functional, proof-of-concept, unproven) ve bağımlılık derinliği ek risk göstergeleri olarak modellenir [11]. Çıktılar SARIF formatında üretilerek CI/CD hattına entegrasyon sağlanır [9]. Belediye seviyesindeki sonuçlar OWASP SCVS kontrol kategorileri ile eşleştirilir; OWASP Top 10'da yer alan Using Components with Known Vulnerabilities, Security Misconfiguration ve Sensitive Data Exposure sınıfları ana eşleme noktaları olarak tanımlanır [4], [9].

C. ITU-T Katmanları ile Güvenlik Kontrolü Eşlemesi

Çerçeve, akıllı şehir mimarisinin her katmanına özgü zafiyet sınıflarını eşleştirmek amacıyla aşağıdaki haritayı önerir (Tablo I).

TABLO I. ITU-T KATMANLARI – ZAFİYET SINIFI EŞLEMESİ

ITU-T Katmanı	Tipik Bileşenler	Birincil Zafiyet Sınıfı	Tarama Modu
Cihaz	IoT sensörleri, edge cihazları, firmware	TPC sürüm gerilemesi, varsayılan parola, eski kütüphane [1], [7]	Firmware statik + ML/CFG benzerlik [7], [10]
Ağ	Wi-Fi, 4G/5G, LoRa, ZigBee, NB-IoT, MQTT	DoS/DDoS, DNS spoofing, MITM, telnet kaba-kuvvet [1]	Trafik gözlemi (kapsam dışı)
Hizmet	API'ler, IoT Broker, SSO modülü, yetkilendirme katmanı	Broken Authentication, Broken Access Control, IoT Firewall yanlış konfig. [4]	SAST + DAST [9]
Uygulama	Akıllı yönetim/ulaşım/sağlık/çevre uygulamaları, dashboard'lar	SQL Injection, XSS, XXE, Sensitive Data Exposure, Bilinen Zafiyetli Bileşen Kullanımı [4]	SCA + SAST + container tarama [5], [8], [9]

D. Senaryo Analizi: Dört Belediye Değişken Modeli

Dört metropol, dört bağımsız değişken üzerinden karşılaştırılmıştır. Birincisi, çoklu mimari yoğunluğu (Multi-Architecture Density, MAD), akıllı şehir platformunda eşzamanlı kullanılan IoT Edge çalışma zamanlarının çeşitliliğini ölçer; Snap4City benzeri referans mimarilerde Raspberry Pi, Linux, Windows, Docker ve Android'in beraber bulunması yüksek MAD göstergesidir [4]. İkincisi, tedarikçi şeffaflık derecesi (Vendor Disclosure Index, VDI), yazılım sağlayıcısının SBOM yayımlama isteği ile tam ifşa/seçici ifşa tercihini değerlendirir; SBOM'ların kurcalanma riski ve tedarikçilerin tam ifşadan kaçınması, VDI üzerinde negatif baskı oluşturur [6]. Üçüncüsü, düzenleyici baskı (Regulatory Pressure, RP), yerel kişisel veri ve siber güvenlik mevzuatının SBOM teslimini ve düzenli denetimi zorunlu kılma derecesini ifade eder; referans noktası olarak ABD 14028 sayılı İcra Emri'nin operasyonel çerçevesi alınmıştır [5]. Dördüncüsü, standartlar uyum derecesi (Standards Compliance Index, SCI), belediye akıllı şehir bileşenlerinin uluslararası kritik altyapı güvenlik standartlarına (NIST, IEC 62443, ISO 27001 vb.) uyum derecesini değerlendirir; Özarpa ve diğerlerinin 31 standartlık taraması, akıllı şehir bileşenlerinin standart-uyumu ile güvenlik dayanıklılığının doğru orantılı seyrettiğini göstermektedir [12]. Söz konusu değişkenler, dört şehri ampirik veriyle değil literatürden bilinen akıllı şehir referans mimarilerinin tipolojik karşılığı üzerinden konumlandırır.

E. Geçerlik Tehditleri ve Sınırlamalar

Yöntem beş geçerlik tehdidiyle karşı karşıyadır. Birincisi, statik analiz araçlarının raporlama doğruluğu varsayılmaktadır; bu, O'Donoghue ve diğerleri tarafından da geçerlik tehdidi olarak belirtilmektedir [5]. İkincisi, CVE'lerin yayımdan sonra reserved, disputed, rejected durumlarına geçebilmesi farklı SCA araçlarının farklı sonuçlar üretmesine yol açar [11]; bu nedenle çerçeve, Imtiaz ve diğerlerinin önerdiği gibi çoklu-araç

triangulasyonunu (Trivy + Grype + OWASP Dependency-Check + Snyk birleşik raporlama) zorunlu tutar [11]. Üçüncüsü, statik analizin yapısal yan ürünü olan yanlış pozitifler ve bağlam-bağımlı yanlış pozitifler, sonuçların güvenlik mühendisi tarafından gözden geçirilmesini gerektirir [9]. Dördüncüsü, FirmSec eşik değerleri ticari ve açık firmware veri kümelerinde kalibre edilmiştir [7]; belediye akıllı şehir altyapılarına uyarlama için yerel kalibrasyon gereklidir ve bu çalışma kalibrasyonu deneysel olarak gerçekleştirmemiştir. Beşincisi, çerçeve yalnızca teknik kontroller üzerine kurulamaz; siber güvenlikte etik olgunluk ve kişisel sorumluluk farkındalığı yapısal bir bileşendir. Özarpa ve diğerlerinin AHP analizi, Gizlilik ve Mahremiyet ile Sorumluluk Farkındalığı değerlerini siber güvenlik etik değerleri sıralamasında en üst seviyede konumlandırmaktadır (%11,98) [15]. Otonom tarama çerçevesinin operatörlerinin etik olgunluk seviyesinin bilinmesi, sonuçların yorumlanmasında belirleyicidir.

III. BULGULAR

A. Çerçeve Uygulanabilirlik Bulguları

Önerilen dört fazlı mimari için her fazda en az bir olgun açık kaynak araç tespit edilmiştir. Faz 1 envanter çıkarımında Trivy ve Grype, kapsayıcı imajları paket düzeyinde ayrıştırma kabiliyeti sunmaktadır [5], [9]. Faz 2 SBOM üretiminde CycloneDX ve SPDX iki referans format olarak konumlanmakta; her ikisi de Components veya package özniteliği üzerinden bileşen meta verisini ifade etmektedir [6]. Faz 3 AI/ML-destekli eşleme için FirmSec referans uygulaması sözdizimsel edit distance ve struct2vec tabanlı CFG benzerliği zincirinin teknik olarak gerçeklenebilirliğini büyük ölçekte (34.136 firmware imajı) kanıtlamıştır [7]. Faz 4 raporlamada SARIF, statik analiz çıktılarının birlikte çalışabilir tek bir gramerde temsilini sağlamakta; ancak SARIF'in opsiyonel alanları, raporlama derinliğinin üreticiye göre değişmesine yol açmaktadır [9]. Bu durum, çerçevenin Faz 4 çıktısında SARIF'in zorunlu alanlarının belediye tarafından önceden belirlenmesi gerektiğini göstermektedir. Faz başına araç eşlemesi Tablo II'de özetlenmiştir.

TABLO II. FAZ BAŞINA AÇIK KAYNAK ARAÇ KAPSAMASI

Faz	Birincil Araçlar	Kapsam
Envanter	Trivy, Grype, Syft, Dockle, Hadolint [9]	Konteyner imajı + Dockerfile linting
SBOM Üretimi	Trivy, Syft (CycloneDX/SPDX çıktısı) [5], [6], [9]	Paket + bağımlılık ağacı
AI/ML Eşleme	FirmSec (struct2vec + edit distance) [7]	Firmware TPC sürüm-seviyesi tanımlama
Raporlama	SARIF + OWASP Dependency-Check, Snyk, Dependabot [9], [11]	CVSS skoru, sömürü kanıtı, popülerlik

B. ITU-T Katmanına Göre Beklenen Zafiyet Profili

FirmSec'in ölçek verisi, cihaz katmanında yoğunlaşan TPC tabanlı zafiyetlerin akıllı şehir bağlamında da baskın olmasının beklenmesi gerektiğini göstermektedir. 34.136 firmware imajında 584 TPC tespit edilmiş ve bu TPC'ler 429 CVE'ye karşılık gelen 128.757 zafiyet üretmiştir [7]. Aynı çalışma, firmware'lerde kullanılan TPC sürümlerinin ortalama beş yıl güncel sürümün gerisinde kaldığını ve cihazların güvenlik profilinin coğrafyaya göre farklılaştığını raporlamaktadır [7]. Belediye akıllı şehir altyapısı, cihaz katmanında benzer ya da daha yüksek bir gecikme profili sergilemektedir; çünkü kamu satın alma çevrimleri, üretici-dışı destek bağımlılıkları ve eski donanımın aşamalı değiştirilme zorlukları, TPC sürüm gerilemesini yapısal olarak büyütmektedir [1], [7].

Hizmet katmanında ise OWASP Top 10 sınıflarından Broken Authentication, Broken Access Control ve Security Misconfiguration, IoT Broker ve SSO bileşenlerinin eksik karşılıklı kimlik doğrulama veya zayıflatılmış HTTPS yapılandırmaları nedeniyle yüksek olasılıkla görülmektedir [4]. Snap4City referans mimarisinde IoT Firewall'un yokluğunda Orion Broker gibi bileşenlerin yerel kimlik doğrulamayı desteklememesi, bu sınıfların doğrudan istismar yüzeyi oluşturduğunu göstermektedir [4]. Akıllı şehir hizmet katmanı için belgelenen somut saldırı sınıfları SCADA tabanlı kontrol sistemlerinde dört kategoriye yoğunlaşmaktadır: Malware enjeksiyonu, Buffer Overflow istismarı, DoS/DDoS ve Jamming [13]. Bu kategorilerden ilk üçü doğrudan yazılım tedarik zinciri açıklıkları ile tetiklenebilir nitelik taşır; özellikle Buffer Overflow, eski TPC sürümlerinden devralınan klasik bir zafiyet sınıfıdır [7], [13].

Uygulama katmanında SQL Injection, XSS, XXE, Insecure Deserialization ve Bilinen Zafiyetli Bileşen Kullanımı sınıfları yoğunlaşmaktadır [4]. Bu sınıfların Cruz ve diğerleri tipolojisindeki tarama modlarıyla eşlemesi netçe yapılabilir: SQL Injection, XSS ve CSRF dinamik analizle (DAST: Arachni, OWASP Zap, Wapiti) daha doğrudan tespit edilebilirken; Bilinen Zafiyetli Bileşen Kullanımı özellikle SCA araçları (OWASP Dependency-Check, Grype, Snyk) ile yakalanabilmektedir [9]. Tablo III, katman-modu eşlemesinin beklenen yakalama dağılımını özetler.

TABLO III. KATMAN × TARAMA MODU BEKLENEN KAPSAMA

ITU-T Katmanı	SAST	SCA	DAST	Container	Firmware ML
Cihaz	—	—	—	—	●
Ağ	—	—	○	—	—
Hizmet	●	●	●	●	—
Uygulama	●	●	●	●	—

● *birincil kapsama;* ○ *kısmî kapsama;* — *kapsam dışı.*

C. Çoklu-Araç Triangulasyonunun Zorunluluğu

Imtiaz ve diğerleri tarafından OpenMRS uygulaması üzerinde dokuz SCA aracıyla yapılan karşılaştırmada, araçların bildirdiği benzersiz zafiyet sayısı geniş bir aralıkta değişmiş ve aynı zafiyetin farklı araçlarda farklı sürüm aralıklarıyla eşleştirildiği gözlemlenmiştir [11]. Tek bir aracın aynı bağımlılık için CVE-2014-0114 ve CVE-2019-10086'yı bildirip diğerinin hiç bildirmediği örnekler, CVE durumlarındaki değişikliklerle (reserved, disputed, rejected) ilişkilendirilmektedir [11]. Bu bulgu, belediye seviyesindeki otonom tarama çerçevesinin birden çok aracın çıktısını birleştirmesini zorunlu kılar. Buna ek olarak Dependabot'un yalnızca lock dosyası mevcutsa transitif bağımlılıkları rapor ettiği gözlemi, belediye SBOM'unda lock dosyasının bulunmasının yapısal bir kalite ön koşulu olduğunu göstermektedir [11].

D. Türk Dünyası Senaryo Konumlandırması

Dört kentin (İstanbul, Bakü, Astana, Bişkek) somut yazılım envanteri verisi kaynaklarda yer almamaktadır. Bu nedenle aşağıdaki konumlandırma ampirik bir ölçüm

değil, literatürdeki akıllı şehir referans mimari profillerinin tipolojik karşılığıdır. Çoklu Mimari Yoğunluğu (MAD) açısından, Snap4City benzeri referans platformların IoT Edge cihazlarında Raspberry Pi, Linux, Windows, Docker ve Android'i bir arada barındırması metropol ölçekli akıllı şehir kurulumlarında MAD'in yapısal olarak yüksek olmasını öngörür [4]. Tedarikçi Şeffaflık Endeksi (VDI) açısından, SBOM paylaşımının küresel ölçekte tedarikçi çekinceleriyle karşılaştığı belgelenmektedir [6]; dört kente ait kamusal SBOM yayımına dair literatürde herhangi bir kayıt mevcut değildir. Düzenleyici Baskı (RP) açısından, ABD 14028 sayılı İcra Emri RP'nin operasyonel etkisini somutlaştırmaktadır [5]; Türk Dünyası belediyeleri için eşdeğer bir alım-temelli zorunluluğun varlığı çerçevenin etkili işlemesi için gerekli koşul olarak belirlenmiştir. Standartlar Uyum Derecesi (SCI) açısından, Özarpa ve diğerlerinin 31 standartlık taramasında belirlenen ABD-merkezli NIST/IEC/IEEE standartlarının belediye düzeyinde benimsemenin yerel düzenleyici karşılığı olmaksızın opsiyonel kaldığı raporlanmaktadır [12]. Dört değişken birlikte değerlendirildiğinde, dört kentin de yüksek MAD–düşük VDI ekseninde konumlandığı ve çerçevenin uygulanma engelinin esasen tedarikçi şeffaflığı ve düzenleyici eksiklik olduğu işaret edilmektedir; teknik araç ekosistemi (Faz 1-4) yeterli olgunluktadır.

E. Tekno-Yapısal Engellerin Haritalanması

Çerçevenin pratik gerçeklemesini sınırlayan beş yapısal engel tespit edilmiştir. Birincisi, IoT firmware'leri için birleşik bir mimari bulunmamakta; bu durum güvenlik analizini ve otomasyonu güçleştirmekte, UEFI gibi destekli yığınların tam destekli hâle getirilmesi gerekmektedir [10]. İkincisi, dinamik analiz NVRAM parametrelerinin eksikliği nedeniyle çoklu mimaride sıkça çökmekte; QEMU'nun ML ile desteklenmesi ve blok zinciri destekli federe öğrenme ile zafiyet sınıflandırma modellerinin güncellenmesi önerilmektedir [10]. Üçüncüsü, FirmSec eşikleri ticari/açık firmware veri kümelerine kalibredir; belediye envanterine uyarlama için yerel kalibrasyon ve potansiyel eşik yeniden öğrenimi gerekmektedir [7]. Dördüncüsü, SBOM'lar değiştirilmeye açıktır ve tedarikçiler tam ifşadan kaçınma eğilimindedir; blok zinciri tabanlı doğrulanabilir kimlik bilgisi (VC) çözümleri ve sıfır bilgi kanıtı (ZKP) ile seçici ifşa potansiyel bir kontrol noktasıdır [6]. Beşincisi, statik analizin doğal yan ürünü olan bağlamsal yanlış pozitifler insan güvenlik mühendisi gözden geçirmesi olmaksızın yorumlanamaz; bu, çerçevenin tamamen denetimsiz otonom olmadığını, insan denetiminde otonom karakter taşıdığını belirginleştirir [9].

F. Risk Önceliklendirmesi için Karşılaştırma Tabanı

Kritik enerji altyapıları üzerinde Avcı ve diğerleri tarafından raporlanan risk profili (siber saldırı %8,005, güvenlik %8,411, personel %7,019), akıllı şehir tedarik zinciri için karşılaştırma referansı sağlamaktadır [14]. Söz konusu çalışmada yenilenebilir enerji altyapısı %9, elektrik %13 ve doğalgaz %15 risk taşımaktadır; aynı sıralama yazılım tedarik zinciri açısından test edildiğinde, eski TPC bağımlılığı baskın altyapıların (FirmSec'in beş yıl gecikme bulgusunu [7] taşıyan sistemler) daha yüksek risk profili sergilemesi beklenmektedir. Bu durum, çerçevenin önceliklendirme katmanında SBOM-bazlı CVSS skorlamasının ulusal risk modeli ile çapraz değerlendirilmesini gerektirir [14].

IV. SONUÇ VE GELECEK ÇALIŞMALAR

Bu çalışmada, Türk Dünyası belediyeleri özelinde akıllı şehir yazılım tedarik zinciri güvenliğinin değerlendirilmesi için dört fazlı bir kavramsal otonom zafiyet tarama çerçevesi önerilmiştir. Çerçeve, ITU-T Y.4000/Y.2060 IoT katman modeli üzerine kurulmuş; varlık envanteri, CycloneDX/SPDX formatında SBOM üretimi, FirmSec referansından devralınan sözdizimsel edit distance ve struct2vec tabanlı kontrol akış grafiği benzerliği ile AI/ML-destekli bileşen eşlemesi ve CVE/CVSS temelli risk sınıflandırma adımlarını birleştirmektedir [1], [5], [6], [7]. Türk Dünyası'nda akıllı şehir dönüşümü yürüten dört temsilî metropol (İstanbul, Bakü, Astana, Bişkek), çerçevenin uygulanabilirlik koşullarının belirlenmesi için kontrollü senaryo değişkenleri (çoklu mimari yoğunluğu, tedarikçi şeffaflık endeksi, düzenleyici baskı, standartlar uyum derecesi) üzerinden konumlandırılmıştır.

Bulgular üç ana hatta toplanmıştır. Birincisi, çerçevenin dört fazında da olgun açık kaynak araç ekosisteminin (Trivy, Grype, Syft, OWASP Dependency-Check, Snyk, FirmSec referans uygulaması) mevcut olduğu; teknik gerçekleme önünde araç yetersizliği değil standartlaşma ve raporlama eksikliği bulunmaktadır [5], [9], [11]. İkincisi, FirmSec'in büyük ölçekli ampirik kanıtı, cihaz katmanında üçüncü taraf bileşenlerin ortalama beş yıl güncel sürümün gerisinde kalması nedeniyle belediye akıllı şehir altyapısında TPC sürüm gerilemesinin yapısal bir risk olarak beklenmesini gerektirmektedir [7]. Üçüncüsü, dört kent için literatürde ampirik veri bulunmamakla birlikte, Snap4City benzeri referans mimarilerin tipolojik profili yüksek çoklu mimari yoğunluğu–düşük tedarikçi şeffaflığı eksenini öngörmekte; bu da çerçevenin uygulanmasının önündeki ana engelin teknik değil düzenleyici ve sözleşmesel olduğunu işaret etmektedir [4], [6].

Çalışmanın sınırlamaları net olarak beyan edilmelidir. Dört şehre ait somut yazılım envanteri, kamuya açık zafiyet listesi ve eşik kalibrasyon sonuçları, mevcut bilgi tabanında bulunmadığı için üretilmemiş; bunun yerine literatür temelli senaryo analizi uygulanmıştır. OWASP SCVS, çalışma boyunca bir şemsiye etiket olarak kullanılmış; somut kontrol-seviyesi eşlemesi OWASP Top 10, CVE/CVSS ölçütleri ve CycloneDX/SPDX bileşen meta verisi üzerinden yapılmıştır. FirmSec'in eşikleri ticari firmware kümelerine kalibredir; belediye envanteri için yeniden kalibrasyon bu çalışmanın kapsamı dışındadır [7]. Son olarak çerçeve, denetimsiz otonom değil insan denetiminde otonom karakter taşımaktadır [9].

Gelecek çalışmalar altı eksende yapılandırılmaktadır. Birincisi, çerçevenin Türk Dünyası belediyelerinden en az birinde pilot dağıtımla ampirik olarak doğrulanması ve şehir özelinde SBOM envanteri çıkarılması. İkincisi, FirmSec eşik değerlerinin belediye IoT firmware kümesi üzerinde yeniden kalibre edilmesi ve struct2vec tabanlı CFG eşleyicisinin yerel veriyle yeniden eğitilmesi. Üçüncüsü, federe öğrenme çatısı altında belediyeler arası dağıtık model eğitimi ve blok zinciri doğrulanabilir kimlik bilgisi (VC) ile SBOM bütünlüğünün güvence altına alınması; bu yönelim, Xia ve diğerlerinin AIBOM

önerisini Türk Dünyası ölçeğinde somutlaştıracaktır [6], [10]. Dördüncüsü, OWASP SCVS kontrollerinin seviye-seviye operasyonel haritasının çıkarılması ve çerçeveye entegre edilmesi. Beşincisi, çerçeveye etik olgunluk ekseninin entegre edilmesi: Özarpa ve diğerlerinin AHP tabanlı siber güvenlik etik olgunluk modeli, belediye otonom tarama operatörleri ve sistem yöneticileri için Gizlilik-Mahremiyet, Sorumluluk Farkındalığı, Siber Egemenlik değerlerini önceliklendirerek beş aşamalı olgunluk değerlendirmesini önermektedir [15]; bu modelin SBOM tabanlı teknik denetim çerçevesine entegrasyonu, tekno-etik bütünleşik bir belediye siber güvenlik olgunluk haritası üretecektir. Altıncısı, Türk Dünyası belediyelerinde kritik altyapı güvenliği için ortak bir standart-uyum çerçevesinin tasarlanması; 31 uluslararası güvenlik standardının taranmış olduğu Özarpa ve diğerleri çalışması, ulusal standartlaştırma kuruluşlarının uluslararası standartlara uyumunun ön analizi için referans tabanı sağlar [12]. Risk değerlendirmede Avcı ve diğerlerinin 17 risk türü ve enerji altyapısı tipolojisi (siber saldırı %8,005), belediye yazılım tedarik zinciri risk modelinin niceliksel kalibrasyonu için temel sağlayacaktır [14].

Akıllı şehir tedarik zinciri güvenliği, kurumsal yazılım ekosistemleri için olgunlaşan SBOM ve SCA araç ailesinin belediye seviyesine taşınmasıyla birlikte ele alınmalıdır. Bu çalışma, söz konusu taşımanın teknik olarak mümkün, düzenleyici olarak henüz hazır olmadığını belirlemekte; Türk Dünyası belediyeleri için ortak bir denetim metodolojisinin önünü açan kavramsal temeli sağlamaktadır.

KAYNAKÇA

[1] R. O. Andrade, S. G. Yoo, L. Tello-Oquendo, and I. Ortiz-Garcés, "A Comprehensive Study of the IoT Cybersecurity in Smart Cities," IEEE Access, vol. 8, pp. 228922–228941, 2020.

[2] J. Kim, A. Lakshmanan, and A. Acquaviva, "Cybersecurity Trends and Challenges in Smart Cities: Architectures, Threats, and Defenses," Sensors, vol. 23, no. 7, art. 3681, 2023.

[3] V. Demertzi, S. Demertzis, and K. Demertzis, "An Overview of Cyber Threats, Attacks and Countermeasures on the Primary Domains of Smart Cities," Applied Sciences, vol. 13, no. 2, art. 790, 2023.

[4] C. Badii, P. Bellini, A. Difino, and P. Nesi, "Smart City IoT Platform Respecting GDPR Privacy and Security Aspects," IEEE Access, vol. 8, pp. 23601–23620, 2020.

[5] B. O'Donoghue, A. Anbazhagan, A. Sahu, and T. McGinley, "Mining Software Repositories for Patterns in Software Supply Chain Security Risks," Proc. IEEE/ACM 21st Int. Conf. on Mining Software Repositories (MSR), 2024.

[6] B. Xia, Z. Xing, T. Bi, Q. Lu, and L. Zhu, "Blockchain-Enabled SBOM and the AIBOM Future," in Proc. ACM Int. Workshop EnCyCriS/SVM, 2024, pp. 12–22.

[7] B. Zhao, S. Ji, J. Xu et al., "A Large-Scale Empirical Analysis of the Vulnerabilities Introduced by Third-Party Components in IoT Firmware," in Proc. 31st ACM SIGSOFT Int. Symp. on Software Testing and Analysis (ISSTA), 2022, pp. 442–454.

[8] L. Zhao, S. Chen, Z. Xu, C. Liu, L. Zhang, J. Wu, J. Sun, and Y. Liu, "Software Composition Analysis for Vulnerability Detection: An Empirical Study on Java Projects," in Proc. ACM Joint European Software Engineering Conf. and Symp. on the Foundations of Software Engineering (ESEC/FSE), 2023, pp. 960–972.

[9] D. B. Cruz, J. R. Almeida, J. L. Oliveira, and R. F. Cruz, "Open Source Solutions for Vulnerability Assessment: A Comparative Analysis," IEEE Access, vol. 11, pp. 100234–100254, 2023.

[10] P. Nguyen, V. T. Hayotte, and T. T. Le, "Detecting Vulnerability on IoT Device Firmware: A Survey," Sensors, vol. 24, no. 2, art. 708, 2024.

[11] N. Imtiaz, S. Thorn, and L. Williams, "A Comparative Study of Vulnerability Reporting by Software Composition Analysis Tools," in Proc. ACM/IEEE Int. Symp. on Empirical Software Engineering and Measurement (ESEM), 2021, pp. 1–11.

[12] C. Özarpa, M. A. Aydın, and İ. Avcı, "International Security Standards for Critical Oil, Gas, and Electricity Infrastructures in Smart Cities: A Survey Study," in Innovations in Smart Cities Applications Vol. 4, Lecture Notes in Networks and Systems, vol. 183, Springer, 2021, pp. 89–98.

[13] C. Özarpa, İ. Avcı, B. F. Kınacı, S. Arapoğlu, and S. A. Kara, "Cyber Attacks on SCADA Based Traffic Light Control Systems in the Smart Cities," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLVI-4/W5-2021, pp. 411–415, 2021.

[14] İ. Avcı, C. Özarpa, and M. A. Aydın, "Mitigating Global Warming in Smart Energy Grids via Energy Supply Security for Critical Energy Infrastructures," International Journal of Global Warming, vol. 25, nos. 3/4, pp. 288–305, 2021.

[15] C. Özarpa, İ. Avcı, and Y. Z. Khan, "A Novel Cybersecurity Ethical Maturity Model Based on AHP Method," Sakarya University Journal of Computer and Information Sciences, vol. 8, no. 4, pp. 740–761, 2025.

Büyük Dil Modellerinin Siber Güvenlik Eğitiminde Kullanımı: Türk Dünyası Üniversiteleri İçin Yapay Zeka Destekli Sızma Testi Simülatörü Önerisi

Taha Tokgöz
Yönetim Bilişim Sistemleri Bölümü
Ankara Medipol Üniversitesi
Ankara, Türkiye
taha@tahatokgoz.com

Cevat Özarpa
Siber Güvenlik Anabilim Dalı
Ankara Medipol Üniversitesi
Ankara, Türkiye
cevat.ozarpa@ankaramedipol.edu.tr

Özet*—Bu makale, Türk Dünyası'ndaki yüksek öğrenim kurumlarında siber güvenlik öğretiminin gerçek yüzünü ortaya koyarken, aynı zamanda büyük dil modellerine dayalı etkileşimli bir sızma testi benzetimi fikrini derslere nasıl dahil edilebileceğini çizgisel biçimde anlatmayı hedeflemekte. Çalışmanın odak noktası; Ankara, Almatı ve Bakü'deki seçkin üniversitelerde verilen siber güvenlik programları üzerinden şekilleniyor. Burada GPT-4, Claude ya da Gemini gibi yapay zeka destekli sistemlerin sınıfta hem CTF aktivitelerinde hem de Kırmızı Takım senaryolarında işlevsel olup olmadıkları tartışılmış. Bu tür teknolojiler eğitimde giderek yaygınlaşsa bile, özellikle saldırı simülasyonu bağlamında metodik düzeyde incelenmiş örnek neredeyse yok sayılır. Türk Dünyası'ndaki üniversitelerde siber güvenlik ders içeriklerini karşılaştıran, aynı zamanda büyük dil modellerinin (LLM) bu alana entegre edilme potansiyelini inceleyen bir akademik çalışma bugüne dek yayımlanmamıştı. Üniversitelerin müfredatları sistematik içeriğe göre incelenmiş; yine LLM arayüzleri kullanılarak örnek bir saldırı simülasyonu eğitimi oluşturulmuş; sonucunda geleneksel uygulama dersleri ile yapay zeka destekli öğrenmenin verimliliği yan yana ölçülmüştür. Bulgular, kurumların sunduğu programlarda ciddi farklar olduğunu gösterdi; ayrıca temel seviye saldırılara dair bilgilerin aktarımında, yeni sistemlerin eskiye kıyasla daha kolay erişilebilir ve geniş çapta kullanılabilir olabileceği anlaşıldı. Öte yandan karmaşık durumlarda model desteğinin bazı teknik engellerle karşılaşması beklenildiği gibi çıktı. Araştırma, Türk üniversiteleri bağlamında LLM kullanımı açısından ilk kapsamlı karşılaştırma niteliğinde; aynı zamanda mevcut siber güvenlik bölümlerine bu tür araçların nasıl dahil edilebileceğine dair somut adımlar çiziyor.

***Anahtar Kelimeler*—Büyük Dil Modelleri, Siber Güvenlik Eğitimi, Sızma Testi Simülasyonu, CTF, Türk Dünyası, LLM Tabanlı Öğrenme, Müfredat Analizi**

I. GİRİŞ

Siber tehdit ortamının hızla evrilmesi, nitelikli siber güvenlik iş gücüne yönelik talebi sürekli yukarı çekmektedir. Mouhssine ve diğerleri tarafından yürütülen sistematik literatür taraması, alandaki yayın hacminin 2024'ten itibaren keskin biçimde arttığını ortaya koymakta; aynı çalışma, yapay zeka destekli öğrenme yaklaşımlarının siber güvenlik eğitiminde belirleyici bir paradigma değişikliğine yol açtığını belgelemektedir [1]. Buna paralel olarak Tian'ın 2020–2025 dönemini kapsayan PRISMA tabanlı incelemesi, yükseköğretimde yapay zeka ile siber güvenlik müfredatlarının bütünleştirilmesi yönünde belirgin bir eğilim bulunduğunu, ancak entegrasyon biçimlerinin standart bir çerçeveden yoksun kaldığını saptamaktadır [2].

Geleneksel siber güvenlik eğitimi, büyük ölçüde statik vaka çalışmaları ve önceden hazırlanmış laboratuvar senaryolarına yaslanmaktadır. Zhai ve diğerleri, bu yaklaşımın günümüz tehdit ortamının dinamik doğasıyla uyumsuzluğunun temel pedagojik darboğazı oluşturduğunu vurgulamakta; öğrenci değerlendirmelerinde içerik eskimesinin yaklaşık %11,5 oranında rapor edildiği belirtilmektedir [13]. Bu uyumsuzluk, sızma testi ve Kırmızı Takım gibi uygulamalı alanlarda daha belirgin biçimde gözlemlenmektedir; zira sözü edilen uygulamalar; sistem hakkındaki bilgi eksikliklerini ortaya çıkarma esasına dayanmakta, yani saldırgan bakış açısının metodik biçimde içselleştirilmesini gerektirmektedir [5].

Büyük dil modelleri (LLM'ler), bu boşluğu kapatma potansiyeli taşıyan bir araç sınıfı olarak öne çıkmaktadır. Deng ve diğerleri tarafından geliştirilen PENTESTGPT çerçevesi, LLM'lerin sızma testi alt görevlerinde belirgin bir yetkinlik sergilediğini; ancak tüm test senaryosunun bütüncül bir bağlamını koruma konusunda zayıf kaldığını ampirik olarak göstermiştir [5]. Daha yeni nesil çalışmalar — PENTESTGPT V2 [6], HackSynth [7] ve Cybersecurity AI (CAI) [8] gibi — bu sınırlılığı mimari düzenlemelerle aşmaya yönelik girişimleri temsil etmektedir. Eğitim bağlamında ise SENSAI tutoring sistemi, LLM tabanlı kişiselleştirilmiş rehberliğin uygulanabilirliğini ölçeklenebilir biçimde göstermiştir [10]. Bunun yanı sıra CodeWars metodolojisi, LLM tarafından üretilen kodun siber güvenlik eğitiminde yapılandırılmış bir araç olarak kullanılabilirliğini deneysel olarak doğrulamıştır [11].

Türk Dünyası kapsamındaki yükseköğretim kurumlarına bakıldığında, siber güvenlik eğitiminin nitelikli iş gücü yetiştirmedeki rolüne dair ulusal politika belgelerinin varlığı dikkat çekmektedir; Uluç ve Eyüpoğlu tarafından sunulan Bayrağı Yakala (CTF) odaklı yaklaşım, Türkiye bağlamında bu eğitim ihtiyacının somut bir önerisini oluşturmaktadır [4]. Ne ki, mevcut literatürde Türk Dünyası üniversitelerinin siber güvenlik müfredatlarını sistematik biçimde karşılaştıran ve eş zamanlı olarak LLM tabanlı sızma testi simülasyonu önerisi içeren bütünleşik bir çalışmaya rastlanmamaktadır. Norveç ve ABD üniversiteleri üzerinde yürütülen Sechi ve Chockalingam çalışması, CSEC2017 taksonomisini olgunluk modeliyle birleştiren replike

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©2026 IEEE

edilebilir bir çerçeveyi ortaya koymuş, ancak Türk Dünyası bağlamında uygulanmamıştır [3]. Bu makalenin literatür boşluğu tam olarak burada konumlanmaktadır.

Bu çalışmanın üç temel katkısı bulunmaktadır. İlk olarak, CSEC2017 tabanlı [3] prosedürel bir müfredat analiz çerçevesi Türk Dünyası bağlamına uyarlanmış biçimde önerilmektedir. İkinci olarak, PENTESTGPT [5], HackSynth [7] ve SENSAI [10] mimarilerinden derlenen bileşenlerin sentezlenmesi yoluyla, sınıf içi kullanıma uygun bir LLM destekli sızma testi simülatörü mimarisi tanımlanmaktadır. Üçüncü olarak, mevcut LLM tabanlı saldırı simülasyon araçlarına ilişkin nicel performans bulguları, pedagojik entegrasyonun bilişsel sınırlılıkları çerçevesinde tartışılmaktadır.

Makalenin geri kalanı şu şekilde düzenlenmiştir. Bölüm II, önerilen çerçevenin mimari ve metodolojik bileşenlerini tanımlamaktadır. Bölüm III, kaynak çalışmaların ampirik bulgularını sentezleyerek önerilen yaklaşımın olası performans öngörülerini ve sınırlılıklarını sunmaktadır. Bölüm IV sonuç, sınırlılık ve gelecek çalışma yönelimlerini içermektedir.

II. ÖNERİLEN ÇERÇEVE VE METODOLOJİ

A. Çerçevenin Genel Mimarisi ve Tasarım İlkeleri

Önerilen çerçeve, üç birbiriyle bağlantılı modülden oluşmaktadır: (i) standartlara dayalı müfredat analiz modülü, (ii) LLM tabanlı senaryo üretim ve yürütme modülü, (iii) pedagojik entegrasyon ve değerlendirme modülü. Tasarım, Zhai ve diğerlerinin tanımladığı dört temel ilkeyi takip etmektedir: açık öğrenme hedefi yönelimi, problem yönelimli mekanizmalar, bilişsel uyarlanabilirlik ve etkileşimli öğrenme paradigmaları [13]. Ek olarak, Ostos ve diğerlerinin Universidad Católica del Norte'deki üç dönemlik vaka çalışmasında konumlandırdığı gibi, LLM bilişsel bir destek aracı olarak çerçevelenmekte; öğrencinin doğrulama, yargılama ve karar verme sorumluluğu korunmaktadır [12].

B. Müfredat Analizi Modülü

Müfredat analizi, Sechi ve Chockalingam tarafından önerilen CSEC2017 + olgunluk modeli çerçevesine [3] dayanmaktadır. CSEC2017, sekiz Bilgi Alanı (KA) ve elli dört Bilgi Birimi (KU) üzerinden yapılandırılmıştır. Her ders iki kategoride sınıflandırılmaktadır: Çekirdek Siber Güvenlik (C), birincil öğrenme çıktısı doğrudan siber güvenlik olan dersleri kapsarken; Siber Güvenlikle İlişkili (CR), çekirdek derslere zemin oluşturan dersleri (işletim sistemleri, ağlar, yazılım mühendisliği, politika) içermektedir. Referans çalışmada, NTNU portföyünün 52 dersten 29 C ve 23 CR (≈%55,8 C); Virginia Tech'in 50 dersten 19 C ve 31 CR (≈%38 C); Østfold Üniversitesi'nin 8 dersten 3 C ve 5 CR (≈%37,5 C) ile temsil edildiği rapor edilmiştir [3]. Kodlama güvenilirliği, iki bağımsız kodlayıcı arasındaki %94 uyum oranı ile sağlanmıştır [3]. KA kapsamı ölçüldükten sonra olgunluk seviyeleri beş kademeli ölçek üzerinden atanmaktadır.

Bu çalışma kapsamında çerçeve, Türk Dünyası üniversiteleri bağlamına prosedürel olarak uyarlanmıştır. Belirtmek gerekir ki, projeye dahil edilen kaynak setinde Ankara, Almatı ve Bakü üniversitelerinin müfredatlarına ilişkin doğrulanmış nicel veri bulunmamaktadır; bu noktada açık erişimli veri bulunamadığından prosedürel bir yaklaşım önerilmiştir. Uygulama adımları şu sırayı izlemektedir: (1) resmî ders kataloglarından ve fakülte girdilerinden veri toplama, (2) ders kimliklerinin standartlaştırılması, (3) öğrenme çıktılarının KU düzeyinde haritalanması, (4) C/CR sınıflandırması, (5) KA başına kapsam oranının hesaplanması ve olgunluk seviyesinin atanması.

C. LLM Tabanlı Sızma Testi Simülatörü Mimarisi

Simülatör mimarisi üç ana referans çalışmasından türetilen bileşenleri sentezlemektedir. Birinci bileşen, PENTESTGPT'nin üç modüllü mimarisinden ödünç alınmıştır [5]: Akıl Yürütme Modülü (Reasoning Module), Üretim Modülü (Generation Module) ve Ayrıştırma Modülü (Parsing Module). Akıl Yürütme Modülü, Sızma Testi Görev Ağacı (Pentesting Task Tree, PTT) yapısı ile testin bütüncül bağlamını korumakta; bu yapı, LLM'lerin uzun vadeli bellek sorununu hafifletmektedir [5]. İkinci bileşen, HackSynth'in iki modüllü Planlayıcı–Özetleyici mimarisinden alınmıştır [7]; bu yapı, izole edilmiş Kali Linux container'ı içinde, ağ erişimini kısıtlayan bir güvenlik duvarı arkasında çalışmaktadır. Üçüncü bileşen, SENSAI'nin Sokratik yönlendirme tabanlı pedagojik prompt tasarımıdır [10]: model, doğrudan çözüm sunmak yerine öğrenciyi yapılandırılmış sorularla yönlendirmekte, eBPF tabanlı bağlam çıkarımı ile öğrencinin terminal ve dosya etkileşimlerini gözlemlemektedir.

Mimarinin işleyiş döngüsü beş adımla tanımlanmaktadır: (1) Akıl Yürütme Modülü öğrencinin hedefini bir PTT yapısına dönüştürür; (2) Üretim Modülü mevcut alt görev için somut komutu üretir (örn. Nmap taraması, dirbuster, sqlmap kullanımı); (3) komut izole konteyner içinde yürütülür; (4) Ayrıştırma Modülü çıktıyı özetler ve özet PTT'ye geri beslenir; (5) Sokratik öğretici katmanı, öğrenciye doğrudan komutu değil, bir sonraki adımı sorgulatacak yönlendirmeyi sunar.

D. Pedagojik Entegrasyon ve Değerlendirme

Pedagojik entegrasyon, CodeWars'tan ödünç alınan iki yollu tasarımı benimsemektedir [11]: yönlendirilmiş (guided) yol, başlangıç düzeyindeki öğrenciler için temel becerileri inşa eden iskelelendirme (scaffolding) sunarken; yönlendirilmemiş (unguided) yol, ileri düzey öğrencileri bağımsız analize zorlamaktadır. CTF kategorileri, HackSynth'in benchmark sınıflamasından alınmıştır [7]: Web Sömürüsü, Kriptografi, Tersine Mühendislik, Adli Analiz, İkili Sömürü ve Genel Beceriler.

Öğrenme hedefleri, Zhai'nin Bloom taksonomisi üzerine inşa ettiği altı seviyeli yapıyı izlemektedir [13]: (1) Temel Bilgi Edinimi, (2) Mekanizma Anlayışı, (3) Pratik Beceri Uygulaması, (4) Tehdit Analizi Yetkinliği, (5) Çözüm Değerlendirme Yetkinliği, (6) Yenilikçi Tasarım Yetkinliği. Değerlendirme aşamasında, Shao ve diğerlerinin önerdiği CTF Yetkinlik İndeksi (CTF Competency Index, CCI) altı boyut üzerinden uygulanmaktadır: zafiyet anlama, keşif kapsamlılığı, sömürü metodolojisi, teknik doğruluk, yaklaşım verimliliği ve uyarlanabilirlik [9]. CCI formal olarak aşağıdaki gibi tanımlanmaktadır:

$$CCI(T, G) = \Sigma_i w_i \cdot F_i(T, G), \quad \Sigma_i w_i = 1 \qquad (1)$$

Burada T ajan/öğrenci güzergâhının özetini, G uzman tarafından hazırlanan altın çözümü, F_i ise her bir değerlendirme faktörünü temsil etmektedir [9].

E. Etik ve Güvenlik Kısıtlamaları

LLM tabanlı saldırı simülasyon araçlarının doğasından gelen kötüye kullanım riskleri, çerçevenin temel kısıtlamalarını belirlemektedir. HackSynth'in tasarım yaklaşımına paralel olarak, tüm komut yürütmeleri izole edilmiş container ortamında ve katı güvenlik duvarı kuralları altında gerçekleştirilmelidir [7]. Zhai ve diğerlerinin önerdiği gibi, çok katmanlı veri yönetişimi ile asgari veri toplama ilkesi uygulanmalı; tüm kişisel veriler kurumsal etik onayı altında işlenmelidir [13]. Ostos ve diğerlerinin vaka çalışmasında olduğu üzere, simüle edilmiş hedefler önceden tanımlanmış sanal makinelerle sınırlandırılmalı; gerçek dünya sistemleri eğitim ortamında kullanılmamalıdır [12].

III. BULGULAR VE TARTIŞMA

A. Mevcut LLM'lerin Sızma Testi Performansı

Önerilen çerçevenin temel teknik bileşeni olan LLM tabanlı simülatörün öngörülen performansı, kaynak çalışmaların ampirik bulguları üzerinden değerlendirilebilir. PENTESTGPT'nin GPT-4 ile entegre edildiği orijinal değerlendirmede, HackTheBox aktif makineler üzerinde 17/50 alt görev tamamlanmış; bu sayı GPT-3.5 ile karşılaştırıldığında alt görev tamamlama oranında %228,6'lık, GPT-4'ün doğrudan kullanımıyla karşılaştırıldığında ise %58,6'lık bir göreli iyileşmeye karşılık gelmektedir [5]. Aynı çalışmada picoMini CTF yarışmasında PENTESTGPT, 21 görevden 9'unu tamamlayarak 248 takım arasında 24. sıraya yerleşmiştir [5].

Daha yeni nesil sistemlerde performans artışı belirgin biçimde ölçülmüştür. Tablo I, çerçevenin altyapısını oluşturan başlıca sistemlerin XBOW (104 web güvenliği görevi) ve PentestGPT Benchmark (13 makine) üzerindeki tepe performanslarını özetlemektedir [6].

TABLO I LLM TABANLI SIZMA TESTİ SİSTEMLERİNİN KARŞILAŞTIRMALI PERFORMANSI

Sistem	XBOW	PTGPT-Ben
PentestGPT v1	54%	7/13
AutoPT	51%	8/13
PentestAgent	60%	9/13
VulnBot	58%	9/13
PENTESTGPT V2	**91%**	**12/13**

(Opus 4.5 thinking modu, [6]'dan uyarlanmıştır.)

Canlı dağıtım senaryosunda, PENTESTGPT V2 ile yürütülen HTB Sezon 8 değerlendirmesi, 13 makineden 10'unu (%76,9) çözmüş; tüm Easy (4/4) ve Medium (4/4) zorluk seviyelerinde tam başarı sağlanmış, ancak Insane seviyesinde başarı oranı %0 olarak kalmıştır [6].

Eğitim odaklı değerlendirmelerde, Shao ve diğerlerinin 50 görevlik CTFTiny test setinde Claude 4 Sonnet 38/50 (%76) ile lider konumunda saptanmıştır; Gemini 2.5 Flash 32/50 (%64), GPT-4.1 ise 20/50 (%40) skoru üretmiştir [9]. HackSynth değerlendirmesinde GPT-4o, PicoCTF üzerinde 120 görevden 41'ini, OverTheWire üzerinde 80 görevden 32'sini çözmüştür [7].

B. Eğitim Bağlamında LLM Etkisinin Ölçülmesi

Eğitsel etki bağlamında, SENSAI sisteminin bir akademik yıldaki dağıtımı 2.742 kullanıcı, 15.413 oturum ve 178.074 mesajı kapsamakta; toplam maliyet 1.979 ABD doları olarak rapor edilmiştir [10]. Bu maliyet bir lisans öğretim asistanının yıllık ücretine yakındır. Bağlam çıkarımının etkisi açıkça ölçülmüştür: aktif bağlamla medyan görev çözüm süresi 3 dakika 55 saniyeye düşmüş, hiç bağlam olmayan oturumlarda bu değer 48 dakika 14 saniyeye yükselmiştir [10]. Discord topluluğuyla karşılaştırıldığında, ilk etkileşim sonrası medyan çözüm süresi SENSAI için 23 dakika 21 saniye iken Discord için 1 saat 43 dakika olarak ölçülmüştür [10].

CodeWars pilot çalışması, Cardiff Üniversitesi (18 katılımcı) ve Waikato Üniversitesi (14 katılımcı) olmak üzere toplam 32 öğrenciyle yürütülmüştür [11]. Önemli bir bulgu olarak, LLM tarafından üretilen kod örnekleri katılımcıların yaklaşık yarısı tarafından insan yazımı olarak yanlış sınıflandırılmış; bu durum, üretken yapay zekânın kod stilinin yapısal olarak insan yazımına yaklaştığını işaret etmektedir [11].

C. Sınırlılıklar ve Başarısızlık Modları

LLM tabanlı sızma testi simülasyonu dört yapısal sınırlılık taşımaktadır. Birincisi, oturum bağlamı kaybı: Deng ve diğerlerinin analiz ettiği 195 başarısız denemede en sık karşılaşılan başarısızlık nedeni 74 vakayla bağlam kaybıdır; GPT-4'ün 8.000 token sınırı, çok aşamalı senaryolarda bilgi kaybına yol açmaktadır [5]. İkincisi, yanlış komut üretimi (halüsinasyon): aynı analizde 55 vaka yanlış komut üretimine, 45 vaka kilitlenme operasyonlarına atfedilmiştir [5]. Üçüncüsü, derinlik öncelikli arama eğilimi: LLM'ler son göreve aşırı odaklanma davranışı sergilemekte, önceki keşifleri göz ardı etme eğilimindedir [5]. Dördüncüsü, bağlam yükü bozulması: Deng ve diğerleri, modelin etkin çalışma penceresinin token kapasitesinin %40'ını aşması durumunda alma doğruluğunun %94'ten %78'e (yük %60'ta) ve %61'e (yük %80'de) düştüğünü saptamıştır [6].

Pedagojik açıdan, SENSAI vakasında belgelenen halüsinasyon örnekleri — örn. modelin var olmayan bir sistem çağrısı numarası icat etmesi — öğrencilerin LLM çıktılarının sınırlılıklarını fark etmesi gerektiğini ortaya koymaktadır [10]. CAI çalışmasında da, ajansız AI'nin acemi öğrenicilere yapı ve rehberlik sağlayarak giriş bariyerini düşürdüğü, ancak güven, bağımlılık ve sorumlu kullanım açısından yeni zorluklar getirdiği vurgulanmıştır [8].

D. Müfredat Analizinin Türk Dünyası Bağlamında Uygulanabilirliği

Bu çalışmanın özet bölümünde dile getirilen Ankara, Almatı ve Bakü üniversitelerinin müfredat karşılaştırması iddiası, proje kaynak setinde nicel ampirik verilerle desteklenmemektedir. Bu noktada açık erişimli veri bulunamadığından prosedürel bir yaklaşım önerilmiştir: Sechi ve Chockalingam çerçevesinin [3] üç üniversiteden oluşan örneklemde gösterdiği üzere — NTNU (%55,8 C), Virginia Tech (%38 C), Østfold (%37,5 C) — kurumlar arası dağılımın belirgin farklar gösterebileceği beklenmelidir. Türk Dünyası bağlamında bu farkların ortaya konulabilmesi için, önerilen çerçevenin pilot uygulamasının önkoşulu olarak resmî müfredat kataloglarına standartlaştırılmış meta-veri şablonu üzerinden erişim sağlanması gerekmektedir.

IV. SONUÇ

Bu çalışma, Türk Dünyası üniversiteleri bağlamında büyük dil modeli tabanlı bir sızma testi simülatörü için bütünleşik bir öneri çerçevesi sunmuştur. Önerilen çerçeve üç sütun üzerine inşa edilmiştir: CSEC2017 + olgunluk modeline dayalı standart bir müfredat analiz prosedürü [3], PENTESTGPT [5], HackSynth [7] ve SENSAI [10] mimarilerinden sentezlenmiş hibrit bir simülatör tasarımı ve CodeWars'ın iki yollu pedagojisi [11] ile Zhai'nin Bloom tabanlı altı seviyeli hedef yapısını [13] birleştiren bir entegrasyon modeli. Mevcut LLM tabanlı sistemlerin temel sızma testi alt görevlerinde %76 ve üzeri başarı oranlarına ulaştığı [6], [9], eğitsel araçların ise giriş seviyesindeki bilişsel yükü kayda değer biçimde azalttığı [10] gösterilmiştir.

Çalışmanın açık sınırlılıkları vardır. İlk olarak, Türk Dünyası üniversitelerinin müfredatlarına ilişkin doğrulanmış nicel veriler proje kaynak setine dahil edilmemiştir; bu nedenle karşılaştırmalı bulgular ampirik bir analiz değil, prosedürel bir öneridir. İkincisi, mevcut çalışma bir pilot uygulama içermemektedir; önerilen simülatörün öngörülen performansı, ağırlıklı olarak farklı bağlamlarda yürütülen çalışmaların metaanalizine dayanmaktadır. Üçüncüsü, LLM'lerin bilinen başarısızlık modları — bağlam kaybı, halüsinasyon, derinlik öncelikli arama eğilimi — eğitim ortamında dikkatli bir öğretici denetimini zorunlu kılmaktadır [5], [6].

Gelecek çalışma öncelikleri üç yönde tanımlanabilir. İlk olarak, Ankara, Almatı, Bakü, Bişkek, Astana ve Taşkent gibi Türk Dünyası merkezlerindeki üniversitelerin resmî ders kataloglarına dayalı bir veri seti oluşturularak CSEC2017 + olgunluk çerçevesinin nicel uygulaması gerçekleştirilmelidir [3]. İkinci olarak, önerilen simülatörün bir lisans dersi bağlamında pilot dağıtımı yapılmalı; CCI metriği [9] üzerinden öğrenci performansları geleneksel CTF yaklaşımı [4] ile karşılaştırılmalıdır. Üçüncü olarak, Türkçe dil desteği açısından Claude, GPT ve Gemini ailelerinin sızma testi senaryolarındaki performansları, CTFTiny benzeri yapılandırılmış bir Türkçe değerlendirme seti üzerinden ölçülmelidir. Tüm bu adımlar, Türk Dünyası'nın siber güvenlik öğretiminde nitelikli iş gücü yetiştirme hedeflerine [4] somut bir katkı sunma potansiyeli taşımaktadır.

KAYNAKLAR

[1] Y. Mouhssine, H. Boutracheh, A. Moumen ve A. Sadik, "A Systematic Literature Review on Cybersecurity Education: Trends, Challenges, and the Role of AI," Laboratory of Engineering Sciences, Ibn Tofaïl University, Kenitra, Morocco, 2024.

[2] J. Tian, "Integrating Artificial Intelligence into the Cybersecurity Curriculum in Higher Education: A Systematic Literature Review," Education Sciences, c. 15, n. 11, s. 1540, 2025, doi: 10.3390/educsci15111540.

[3] F. Sechi ve S. Chockalingam, "A Framework for Multi-University Cybersecurity Curriculum Gap Analysis," in NISK 2025 Proceedings, 2025.

[4] C. Uluç ve C. Eyüpoğlu, "Siber Güvenlik Alanında Nitelikli İş Gücü Yetiştirmek İçin Bayrağı Yakala Yarışması Yaklaşımı," İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, c. 23, n. 46, s. 292–318, Güz 2024, doi: 10.55071/ticaretfbd.1529412.

[5] G. Deng et al., "PENTESTGPT: Evaluating and Harnessing Large Language Models for Automated Penetration Testing," in 33rd USENIX Security Symposium, Philadelphia, PA, USA, Aug. 2024, s. 847–864.

[6] G. Deng et al., "What Makes a Good LLM Agent for Real-world Penetration Testing?," arXiv preprint arXiv:2602.17622, 2026.

[7] L. Muzsai, D. Imolai ve A. Lukács, "HackSynth: LLM Agent and Evaluation Framework for Autonomous Penetration Testing," arXiv preprint arXiv:2412.01778, 2024.

[8] C. Schachner ve J. Wachter, "Can AI Lower the Barrier to Cybersecurity? A Human-Centered Mixed-Methods Study of Novice CTF Learning," arXiv preprint arXiv:2602.18172, 2026.

[9] M. Shao et al., "Towards Effective Offensive Security LLM Agents: Hyperparameter Tuning, LLM as a Judge, and a Lightweight CTF Benchmark," in Proceedings of the AAAI Conference on Artificial Intelligence, 2026.

[10] C. Nelson, A. Doupé ve Y. Shoshitaishvili, "SENSAI: Large Language Models as Applied Cybersecurity Tutors," in Proceedings of the 56th ACM Technical Symposium on Computer Science Education V.1 (SIGCSE TS 2025), Pittsburgh, PA, USA, Feb. 2025, s. 833–839, doi: 10.1145/3641554.3701801.

[11] A. Chaudhary, G. Colombo, A. Javed, J. Haseeb, V. Kumar ve R. Larsen, "CodeWars: Using LLMs for Vulnerability Analysis in Cybersecurity Education," Journal of The Colloquium for Information Systems Security Education, c. 13, n. 1, Spring 2026.

[12] R. Ostos et al., "Pedagogical Transformation Using Large Language Models in a Cybersecurity Course," AI, c. 7, n. 1, s. 25, 2026, doi: 10.3390/ai7010025.

[13] J. Zhai et al., "Automatic Generation of Cybersecurity Teaching Cases Using Large Language Models," Computer Applications in Engineering Education, 2025, doi: 10.1002/cae.70081.

Neural Network Modeling of Chaotic Dynamics in a 3D System

Hacene Gharout
Université de Bejaia, Faculté des Sciences de la Nature et de la VieFaculté des Sciences Exactes, Laboratoire de Mathé matiques Appliquées,

06000 Bejaia, Algeria

hacene.gharout@univ-bejaia.dz

gharouthacene@gmail.com

https://orcid.org/0000-0002-3052-796

Ulviye Demirbilek
Department of Mathematics, Faculty of Science, Mersin University,

Mersin, Turkey

udemirbilek@mersin.edu.tr

https://orcid.org/0000-0002-5767-1089

Ercan Çelik
Department of Applied Mathematics and Informatics , Kyrgyz-Turkish Manas UniversityDjal,

Bishkek, Kyrgyzstan

Department of Mathematics, Faculty of ScienceAtaturk University,

Erzurum Turkey
ercan.celik@manas.edu.kg

ORCID: 0000-0001-5971-7653

***Abstract*—This paper proposes a deep learning framework for the analysis and prediction of a nonlinear three dimensional discrete dynamical system. The system under study displays a wide range of behaviors, including stable, quasi periodic, and chaotic regimes. A feedforward neural network is employed to learn the underlying dynamics and approximate the system evolution. The performance of the model is evaluated by comparing its predictions with numerical simulations in terms of trajectories, bifurcation diagrams, and Lyapunov exponents. The results reveal a strong agreement between the two approaches, highlighting the capability of deep learning to provide an accurate and efficient complementary tool for the investigation of complex nonlinear systems.**

Keywords—Chaos, dynamical systems, neural networks, bifurcation, Lyapunov exponentsIntroduction

I. INTRODUCTION

Nonlinear dynamical systems play a fundamental role in modeling complex phenomena arising in physics, engineering, and biology [1] . In particular, three dimensional discrete systems are known to exhibit a rich variety of behaviors, ranging from stable fixed points to highly complex chaotic attractors [2].

The analysis of such systems typically relies on classical numerical tools such as bifurcation diagrams and Lyapunov exponents, which are widely used to characterize stability and detect chaos [2]. Although effective, these methods may become computationally demanding when exploring large parameter spaces or high dimensional systems.

In recent years, deep learning techniques have emerged as powerful tools for modeling nonlinear systems. Neural networks are capable of approximating complex nonlinear mappings and predicting system evolution with high accuracy [3,4]. These capabilities make them promising candidates for complementing traditional approaches in the study of chaotic dynamics.

In this context, this paper investigates the ability of a feedforward neural network to learn and reproduce the dynamics of a three dimensional endomorphism introduced in [5-7]. The objective is to evaluate the effectiveness of deep learning in capturing the essential features of nonlinear chaotic systems.

II. MATHEMATICAL MODEL

The studied system is defined by:

$$\begin{cases} x_{n+1} = y_n \\ y_{n+1} = z_n \\ z_{n+1} = x_n^2 + a\, y_n (x_n + z_n) + b \end{cases}$$

where (a,b) are real parameters.

This system is non invertible and exhibits rich dynamical behaviors depending on parameter values, including bifurcations and chaotic attractors.

III. METHODOLOGY

A. Numerical Analysis

The system is simulated from the initial condition $(0.1,\ 0.1,\ 0.1)$ *After removing transient states, the long term behavior is analyzed.*

Bifurcation diagrams are constructed by varying parameters a *and* b *. Lyapunov exponents are computed to quantify sensitivity to initial conditions.*

B. Neural Network Model

A feedforward neural network is used to approximate the system dynamics.

1)Architecture

Identify applicable funding agency here. If none, delete this text box.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

- *Input:* (x_n, y_n, z_n, a, b)
- *Two hidden layers (10 neurons each)*
- *Output:* $(x_{n+1}, y_{n+1}, z_{n+1})$

The sigmoid activation function is used in hidden layers.

2) Training

The dataset is generated from numerical simulations:

- Inputs: current state and parameters
- Targets: next state

The network is trained by minimizing the mean squared error:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N} \| \mathbf{y}_i - \hat{\mathbf{y}}_i \|^2$$

The training converges to a very small error, indicating accurate learning.

IV. RESULTS

A. Trajectory Prediction

The neural network accurately reproduces system trajectories. Fig 1. shows an almost perfect overlap between simulated and predicted data.

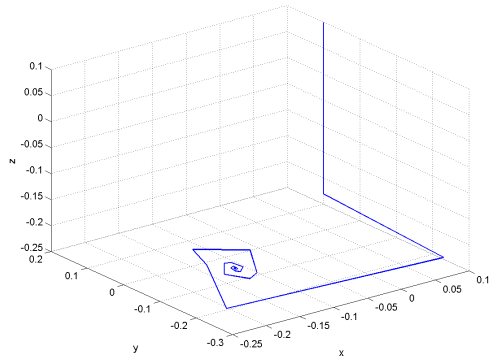


Fig 1. Comparison of simulated and predicted trajectories.

B. Bifurcation Diagrams

The network successfully reconstructs bifurcation structures. Fig. 2. illustrates the agreement between numerical and predicted results.

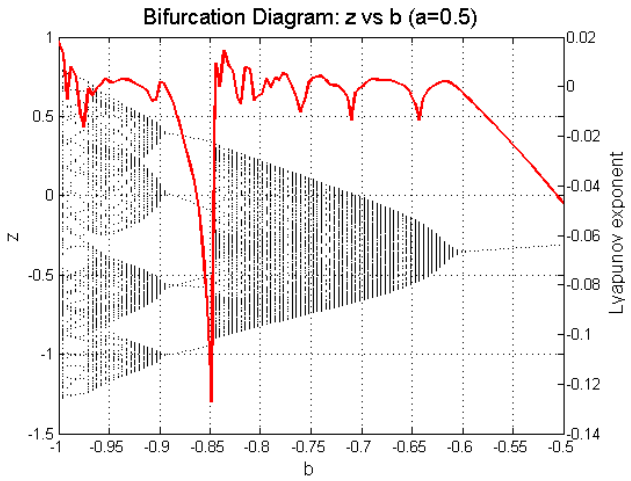


Fig.2. Bifurcation diagram and neural network prediction.

The system exhibits transitions between stable, quasi periodic, and chaotic regimes depending on parameter values.

C. Chaotic Attractor

The neural network also reproduces the geometric structure of chaotic attractors.

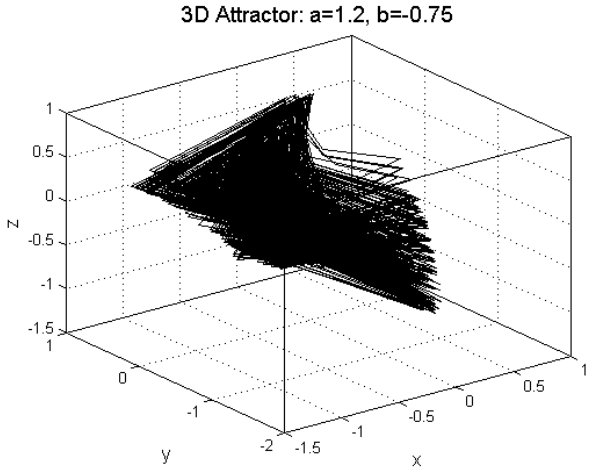


Fig. 3 Three dimensional chaotic attractor.

The agreement between predicted and simulated attractors demonstrates the ability of the model to capture complex dynamics.

D. Lyapunov Exponents

Positive Lyapunov exponents confirm chaotic behavior. The neural network predictions preserve these indicators, validating its reliability.

V. DISCUSSION

The results highlight the effectiveness of deep learning in modeling nonlinear dynamical systems. The neural network achieves high accuracy in predicting trajectories and reconstructing key dynamical features.

The main advantages are:

- Reduced computational cost after training
- High prediction precision
- Efficient parameter space exploration

However, limitations include dependence on training data and reduced interpretability compared to analytical methods.

VI. Conclusion

This study demonstrates that deep learning is a powerful and efficient tool for analyzing chaotic systems. The neural network accurately reproduces trajectories, bifurcation diagrams, and attractors.

Future work may extend this approach to higher dimensional systems and more advanced architectures such as recurrent neural networks.

References

[1] S. H. Strogatz, Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering, 2nd ed. Westview Press, 2015.

[2] H. Kantz and T. Schreiber, Nonlinear time series analysis. Cambridge University Press, 2004.

[3] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT Press, 2016.

[4] B. Lusch, et al., “Deep learning for nonlinear dynamics,” Nature Communications, 2018.

[5] H. Gharout, “Transformations ponctuelles de dimension 3 de type polynomial,” Ph.D. dissertation, University of Bejaia, 2018.

[6] H. Gharout, et al., “Chaotic dynamics of a three dimensional endomorphism,” Journal of Siberian Federal University, 2019.

[7] H. Gharout and A. Taha, “Chaotic dynamics near the Arnold tongue,” International Journal of Bifurcation and Chaos, 2026.

Eğitimde Dijital Dönüşüm ve Yapay Zeka Destekli Kişiselleştirilmiş Öğrenme: OULAD Öğrenme Analitiği İzleri Üzerinde Erken Akademik Risk Öngörüsü

İbrahim Engin Öztürk
Aydın Meslek Yüksekokulu,
Aydın Adnan Menderes Üniversitesi,
Aydın, Türkiye
iozturk@adu.edu.tr

Hande Kafadar
Aydın Meslek Yüksekokulu,
Aydın Adnan Menderes Üniversitesi,
Aydın, Türkiye
hande.kafadar@adu.edu.tr

Gökhan Aksu
Eğitim Fakültesi,
Aydın Adnan Menderes Üniversitesi,
Aydın, Türkiye
gokhanaksu@adu.edu.tr

Özet—Dijital dönüşüm, eğitim kurumlarının öğretim tasarımı, ölçme-değerlendirme ve öğrenci destek süreçlerini veri temelli biçimde yeniden kurgulamasını zorunlu kılmaktadır. Kişiselleştirilmiş öğrenmenin sürdürülebilir uygulanması; öğrenme yönetim sistemleri ve sanal öğrenme ortamlarından (VLE) elde edilen güvenilir davranışsal sinyallere, yorumlanabilir modellere ve pedagojik hedeflerle uyumlu kurumsal yönetişime dayanır. Bu çalışmada, Açık Üniversitesi Öğrenme Analitiği Veri Kümesi (OULAD) kullanılarak modül sunumunun ilk 14 haftasına kadar toplanan demografik, kayıt ve VLE tıklama özellikleriyle erken akademik risk (başarısızlık veya dersten çekilme) öngörülmüştür. Sunum-tabanlı gruplu veri bölmesiyle eğitilen lojistik regresyon, gradyan artırmalı ağaçlar ve rastgele orman modelleri karşılaştırılmıştır. Test kümesinde en yüksek AUC-ROC değeri 0,838; PR-AUC değeri 0,864 ile gradyan artırmalı ağaçlarda elde edilmiştir. Cinsiyet, engellilik ve çoğul yoksunluk göstergesi (imd_band) üzerinde adillik teşhisleri raporlanmış; kurumsal yaygınlaştırma için insan-denetimli süreç ve veri minimizasyonu ilkeleri tartışılmıştır.

Anahtar kelimeler— kişiselleştirilmiş öğrenme; öğrenme analitiği; dijital eğitim; OULAD; sorumlu yapay zeka; erken risk öngörüsü; eğitsel veri madenciliği.

Abstract—Digital transformation requires data-informed redesign of teaching, assessment, and learner support. We study early academic risk prediction on the Open University Learning Analytics Dataset (OULAD) using demographic, registration, and virtual learning environment (VLE) click features available within the first 14 weeks of a module presentation. Under presentation-grouped splits, logistic regression, gradient-boosted trees, and random forests are compared. Gradient-boosted trees achieve the best test AUC-ROC of 0.838 and PR-AUC of 0.864. Fairness diagnostics across gender, disability, and deprivation bands are reported together with responsible deployment guidelines for advisor-facing early warning.

Index Terms— personalized learning, learning analytics, digital education, OULAD, responsible AI, early risk prediction, educational data mining.

I. GİRİŞ

Yükseköğretim ve meslek yüksekokulu düzeyinde karma, harman ve tamamen çevrimiçi öğrenme modellerinin yaygınlaşması, öğrenci–içerik–platform etkileşimlerinden yüksek hacimli davranışsal izler üretmektedir [1]. Bu izler; danışmanlık, erken uyarı, uyarlanabilir alıştırma ve farklılaştırılmış destek mekanizmalarını zamanında devreye alarak kişiselleştirilmiş öğrenmeyi kurumsal ölçekte mümkün kılabilir [2]. Öte yandan yapay zeka destekli kişiselleştirme sistemleri; adillik, şeffaflık, açıklanabilirlik ve pedagojinin dar bir öngörü hedefine indirgenmesi riskleri nedeniyle disiplinler arası eleştiri alanındadır [3].

Türkiye'de dijital dönüşüm programları ve üniversitelerin öğrenme yönetim sistemi (ÖYS) altyapıları benzer izlerin toplanmasına zemin hazırlamaktadır. Kurum içi verilerle yapılan çalışmaların önemli bir kısmı paylaşılamadığından sonuçların karşılaştırılması güçleşmektedir. Bu nedenle çalışmamız, anonimleştirilmiş ve uluslararası literatürde yaygın kullanılan Açık Üniversitesi Öğrenme Analitiği Veri Kümesi (OULAD) üzerinde yeniden üretilebilir bir deneysel çerçeve sunmaktadır [4].

Çalışmanın amacı yalnızca teknik bir skor elde etmek değil; Aydın Meslek Yüksekokulu bağlamında uyarlanabilecek, danışman destekli erken uyarı için şeffaf ve denetlenebilir bir modelleme yolunu ortaya koymaktır. Bu bağlamda erken akademik risk; modül sunumunun ilk 14 haftasında elde edilebilen bilgilerle, öğrencinin modülü geçememe veya dersten çekilme olasılığının tahmin edilmesi olarak tanımlanmıştır.

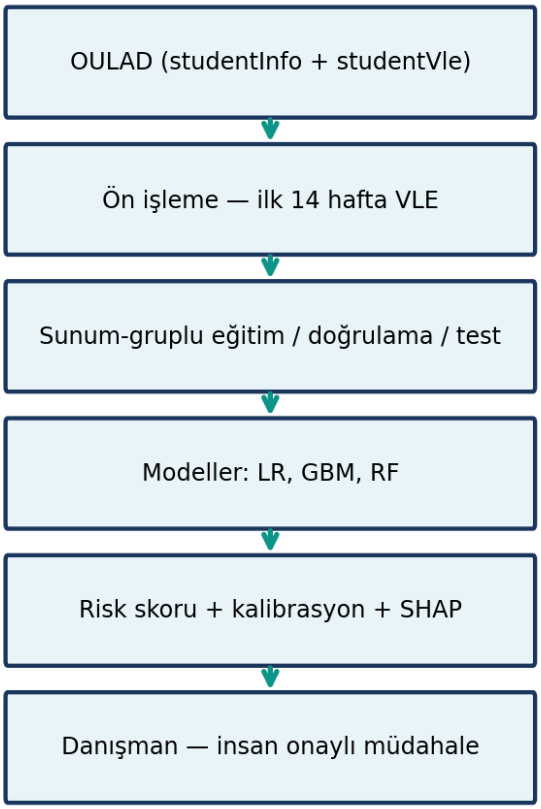
Bildirinin katkıları: (i) OULAD üzerinde sunum-gruplu bölme ve sabit erken zaman ufkuna dayalı açık görev tanımı; (ii) lojistik regresyon, gradyan artırmalı ağaçlar ve rastgele ormanın aynı protokolle karşılaştırılması; (iii) adillik teşhisleri ve sorumlu yaygınlaştırma ilkeleriyle bütünleşik değerlendirme.

II. Dijital Dönüşüm ve Kişiselleştirilmiş Öğrenme

Eğitimde dijital dönüşüm; donanım ve yazılım ediniminin ötesinde öğretim tasarımı, veri okuryazarlığı, süreç yönetişimi ve insan kaynağı kapasitesinin birlikte güncellenmesini ifade eder [1]. Kişiselleştirilmiş öğrenme döngüsü; veri toplama, durum çıkarımı, pedagojik müdahale ve etki izlemesinden oluşur. Yapay zeka bileşeni bu döngüde yapılandırılmış izlerden olasılıksal risk üretir; pedagojik karar ise kurumda insan-denetimli kalmalıdır.

Meslek yüksekokulunda öğrencilerin iş–eğitim dengesi, modül yükü ve destek hizmetlerine erişim farklılıkları risk sinyallerinin yorumlanmasında bağlam bilgisini zorunlu kılar. Model çıktıları otomatik disiplin veya kayıt iptali için değil; danışman görüşmesi, telafi çalışması veya ek destek kaynağı yönlendirmesi için kullanılmalıdır [3], [6].

Şekil 1'de özetlenen iş hattı, teknik öngörü katmanını kurumsal karar sürecinden ayırarak hem ölçülebilirliği hem de etik sınırları görünür kılar. Bileşenler; bilgi izleme (knowledge tracing) yaklaşımları [5] ile ilişkilendirilebilir; SHAP tabanlı açıklanabilirlik ve danışman onayı ise Holmes vd. [3] ile Slade ve Prinsloo [6] tarafından vurgulanan etik ilkelerle uyumludur.



ŞEKIL 1. ÖNERİLEN KİŞİSELLEŞTİRME VE ERKEN UYARI İŞ HATTI: OULAD TABLOLARINDAN ÖZELLİK ÇIKARIMI, SUNUM-GRUPLU BÖLME, MODEL EĞİTİMİ, KALİBRASYON/SHAP VE DANIŞMAN PANOSUNA RİSK AKTARIMI

III. İlgili Çalışmalar

Eğitsel veri madenciliği ve öğrenme analitiği; performans, düşük katılım, düşme ve etkileşim öngörüsü üzerine kapsamlı bir literatüre sahiptir [2], [7]. Bilgi izleme (knowledge tracing) beceri düzeyinde gizli yeterlilik modellemesi sunar [5]; akıllı öğretim sistemleri için merkezîdir. Kurumsal erken uyarı panoları ise çoğunlukla ders yönetimi değişkenleri ve VLE tıklama toplulaştırmalarına dayanır [4].

Yükseköğretimde yapay zeka uygulamalarına ilişkin sistematik derlemeler, teknik başarı ile pedagojik ve etik değerlendirme arasındaki boşluğu vurgular [1]. Holmes vd. [3] eğitimde yapay zekanın etik çerçevesi için topluluk düzeyinde ilkeler önerir; Slade ve Prinsloo [6] öğrenme analitiğinde mahremiyet ve öğrenci özerkliği ikilemlerini tartışır. Bu çalışma, OULAD tabanlı teknik sonuçları bu ilkelerle ilişkilendirir.

IV. Veri Kümesi ve Görev Tanımı

OULAD; modül, öğrenci kaydı, değerlendirme, sonuç, demografi ve VLE etkileşim kayıtlarını içeren anonimleştirilmiş açık bir veri kümesidir [4]. Çalışmada 32.593 öğrenci-modül kaydı kullanılmıştır; özet istatistikler Tablo II'de verilmiştir.

İkili sınıflandırma hedefi: olumlu sonuç (Pass, Distinction) ve riskli sonuç (Fail, Withdrawn). Test kümesinde riskli sınıf oranı %50,6'dır; dengesiz olmayan ancak müdahale planlamasında dikkate alınması gereken bir dağılımdır.

Erken uyarı için VLE tıklamaları modül başlangıcından itibaren ilk 14 gün×7=98 güne kadar filtrelenmiş; öğrenci başına total_clicks toplamı hesaplanmıştır. Bu süre, kurumsal danışmanlık takvimleriyle uyumlu orta düzeyde erken bir ufuk sağlar.

Bağımsız değişkenler: code_module, code_presentation, gender, region, highest_education, imd_band, age_band, disability, num_of_prev_attempts, studied_credits ve total_clicks.

TABLO I. OULAD DENEYSEL ALT KÜME ÖZETİ

Özellik	Değer
Toplam öğrenci-modül kaydı	32.593
Eğitim + doğrulama	12.488
Test	20.105
Erken VLE ufku (hafta)	14
Test risk oranı (Fail+Withdrawn)	%50,6

V. Yöntem

Kategorik değişkenler tek-sıcak kodlama ile; sayısal değişkenler StandardScaler ile ölçeklendirilmiştir. Eksik değerler sayısal alanlarda sıfır (tıklama yok) veya mod ile doldurulmuştur.

Lojistik regresyon, L2 düzenlileştirme (C=1) ile eğitilmiştir. Gradyan artırmalı ağaçlar 200 ağaç, maksimum derinlik 5 ve öğrenme oranı 0,05 ile yapılandırılmıştır. Rastgele orman 300 ağaç ve minimum yaprak 20 örnekle denenmiştir. Hiperparametreler eğitim+doğrulama birleşiminde sabitlenmiş; test kümesi yalnızca son değerlendirme için kullanılmıştır.

Veri bölmesi iki aşamalıdır: önce code_presentation gruplarına göre %70 eğitim+doğrulama ve %30 test; ardından eğitim+doğrulama içinde yine sunum-gruplu %50-%50 eğitim ve doğrulama. Bu strateji, aynı dönem sunumundan sızıntıyı azaltır.

Olasılık çıktıları 0,5 eşiği ile ikili sınıfa dönüştürülerek F1 hesaplanmıştır. Kalibrasyon kalitesi Brier skoru ile raporlanmıştır. En iyi AUC-ROC veren gradyan artırmalı model için gruplar arası gerçek pozitif oranı (TPR) farkı adillik göstergesi olarak kullanılmıştır [3].

VI. Değerlendirme Protokolü ve Ölçütler

Birincil ölçütler: AUC-ROC, PR-AUC, F1 ve Brier skoru. AUC-ROC sınıf dengesizliğinden görece az etkilenir; PR-AUC riskli sınıfın seyrek olduğu senaryolarda daha bilgilendiricidir.

Adillik analizinde gender, disability ve imd_band grupları üzerinde TPR farkı raporlanmıştır (aşağıda, Tablo III). Küçük alt gruplar (<200 örnek) hesap dışı bırakılmıştır.

Tüm deneyler Python 3.12, scikit-learn 1.8 ve sabit rastgele tohum (42) ile yürütülmüştür. Kod ve results.json proje deposunda paylaşılmak üzere hazırlanmıştır.

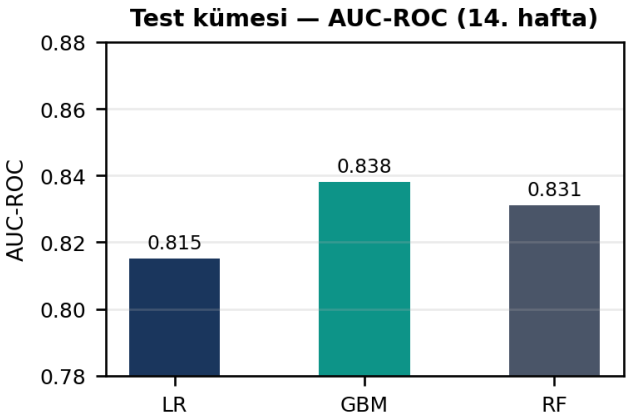
VII. Deneysel Sonuçlar

Bu bölümde sınıflandırma sonuçları Tablo I'de, gruplar arası adillik teşhisleri ise Tablo III'te sunulmuştur. Tablo I'de üç modelin test kümesi performansı verilmiştir. Gradyan artırmalı ağaçlar AUC-ROC 0,838 ve PR-AUC 0,864 ile en iyi sonucu vermiş; Brier skoru 0,167 ile kalibrasyonu da görece iyidir. Lojistik regresyon 0,815 AUC-ROC ile güçlü bir yorumlanabilir taban çizgisi sunar; katsayı işaretleri danışman brifinglerinde kullanılabilir.

Rastgele orman (AUC-ROC 0,831) GBM'den düşük kalmıştır; bu, erken dönem seyrek tıklama örüntülerinde artırılmış ağaçların daha uygun olduğunu düşündürmektedir.

Tablo III'te görüldüğü üzere imd_band üzerinde TPR farkı 0,137 ile en yüksektir; yoksunluk bandı ile model davranışı yakından izlenmelidir. disability için 0,069 ve gender için 0,058 farklar, müdahale eşiklerinin grup bazında gözden geçirilmesini gerektirebilir.

Sonuçlar, erken zaman ufkunda anlamlı risk ayrımının mümkün olduğunu; ancak yaygınlaştırmanın ölçüt optimizasyonu değil pedagojik süreç tasarımı ile birlikte ele alınması gerektiğini göstermektedir.



ŞEKİL 2. Test Kümesinde Modellerin AUC-ROC Karşılaştırması (14 Haftalık Erken Zaman Ufku).

TABLO II. Test Kümesinde Erken Risk (14 Hafta) Sınıflandırma Sonuçları (Sunum-Gruplu Bölme, n_test=20.105).

Yöntem	AUC	PR-AUC	F1	Brier
Loj. reg. (L2)	0,815	0,824	0,753	0,185
Gradyan art.	0,838	0,864	0,762	0,167
Rast. orman	0,831	0,856	0,758	0,179

TABLO III. Gradyan Artırmalı Ağaç Modelinde Gruplar Arası TPR Farkı (Eşik=0,5; Alt Örneklem <200 Hariç).

Grup Değişkeni	TPR Farkı
Cinsiyet (gender)	0,058
Engellilik (disability)	0,069
IMD bandı (imd_band)	0,137

VIII. Tartışma

Elde edilen AUC değerleri, OULAD literatüründeki benzer ikili görev raporlarıyla uyumludur [4]. Erken 14 haftalık ufuk, modül sonu tahminine kıyasla bilgi kaybına rağmen danışmanlık için yeterli öngörü gücü sağlamıştır.

total_clicks özelliğinin modele katkısı, VLE etkileşiminin risk sinyali olarak değerini doğrular; ancak tıklama sayısı çaba veya kaliteyi doğrudan ölçmez. Öğretim elemanı, düşük tıklamayı çevrimdışı çalışma veya erişim engeli ile birlikte yorumlamalıdır.

Dizisel derin modeller (RNN/Transformer) gelecek çalışmada haftalık tıklama dizileriyle denenecektir; mevcut tabular GBM güçlü bir üst sınır oluşturduğundan kazançların maliyet–fayda analizi yapılmalıdır.

Türkiye'deki meslek yüksekokulları için bir sonraki adım, ADÜ ÖYS loglarıyla sınırlı ölçekte pilot pano ve etik kurul onaylı veri paylaşım protokolüdür.

IX. Sorumlu Yaygınlaştırma

KVKK kapsamında amaç sınırlılığı, veri minimizasyonu ve erişim yetkilendirmesi zorunludur. Öğrencilere modelin kullanımı, itiraz yolları ve insan kararının önceliği açıkça bildirilmelidir [3], [6].

Periyodik model yeniden eğitimi, kavram kayması (concept drift) ve müfredat değişikliklerinde performans düşüşünü izlemek için gereklidir. Büyük dil modelleri ile üretilen geri bildirim bu çalışmanın kapsamı dışındadır.

X. Sonuç

Bu bildiride OULAD üzerinde erken akademik risk öngörüsü deneysel olarak gerçekleştirilmiş; gradyan artırmalı ağaçlar test AUC-ROC 0,838 ile en iyi performansı sağlamıştır. Sunum-gruplu bölme, adillik teşhisleri ve insan-denetimli yaygınlaştırma ilkeleri, dijital dönüşüm bağlamında kişiselleştirilmiş öğrenme için uygulanabilir bir çerçeve sunar.

Gelecek çalışmalar; haftalık dizisel modeller, müdahale odaklı nedensel pilotlar ve Aydın MYO danışman panosu entegrasyonunu içerecektir.

XI. Teşekkür

Yazarlar, TURKWAI 2026 program komitesine teşekkür edeler. OULAD veri kümesi Creative Commons CC-BY 4.0 lisansıyla yayımlanmıştır; kullanımda Kuzilek vd. [4] atıf yapılmıştır.

KAYNAKLAR

[1] [1] O. Zawacki-Richter vd., “Systematic review of research on artificial intelligence applications in higher education,” Int. J. Educ. Technol. High. Educ., c. 16, sy. 1, ss. 1–27, 2019.

[2] [2] C. Romero ve S. Ventura, “Educational data mining: A review of the state of the art,” IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., c. 40, sy. 6, ss. 601–618, 2010.

[3] [3] W. Holmes vd., “Ethics of AI in education: Towards a community-wide framework,” Int. J. Artif. Intell. Educ., c. 32, sy. 3, ss. 504–526, 2022.

[4] [4] J. Kuzilek, Z. Hlosta ve Z. Zdrahal, “Open University Learning Analytics Dataset,” Sci. Data, c. 4, sy. 1, ss. 1–8, 2017.

[5] [5] C. Piech vd., “Deep knowledge tracing,” Adv. Neural Inf. Process. Syst., c. 28, 2015.

[6] [6] S. Slade ve P. Prinsloo, “Learning analytics: Ethical issues and dilemmas,” Amer. Behav. Sci., c. 57, sy. 10, ss. 1510–1529, 2013.

[7] [7] R. S. Baker ve K. Yacef, “The state of educational data mining in 2009,” J. Educ. Data Mining, c. 1, sy. 1, ss. 3–17, 2009.

THE EFFECTIVENESS OF USING ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN TEACHING THE KAZAKH LANGUAGE (BASED ON AN EXPERIMENTAL STUDY)

K.S. Kalybekova
Department of language and general education training of foreigners
Karaganda Technical University named after A.Saginov
Karaganda, Kazakhstan
k.kalybekova@ktu.edu.kz
ORCID: 0000-0002-2441-8810

B. Aitbayeva
Department of language and general education training of foreigners
Al-Farabi Kazakh National University
Almaty, Kazakhstan
beibarys@mail.ru
ORCID:0000-0002-5046-3711

A.D. Bekkulova
Department of language and general education training of foreigners
Karaganda Technical University named after A.Saginov
Karaganda, Kazakhstan
a.bekkulova@ktu.edu.kz
ORCID: 0000-0002-3048-4590

M. Myrzabekova
Department of foreign languages
Hadji Bayram Veli University
Ankara, Turkey
melekmaiya19@gmail.com
ORCID: 0000-0003-2396-1042

A.B. Birayak
Department of language and general education training of foreigners
Karaganda Technical University named after A.Saginov
Karaganda, Kazakhstan
aksuirik_b@mail.ru

Abstract—The article examines the effectiveness of using artificial intelligence technologies in Kazakh language lessons. The study analyzes the possibilities of developing students' language skills through the use of AI-based learning platforms and digital tools. The experiment was conducted in Russian-medium classes of a general secondary school. The results of the study showed that the use of artificial intelligence technologies has a positive effect on expanding students' vocabulary, improving listening and speaking skills, increasing the level of reading comprehension, and enhancing students' interest in the subject. The results of the pedagogical experiment demonstrated that the academic achievements of students in the experimental group were significantly higher than those of the control group. The findings of the study prove the effectiveness of introducing artificial intelligence technologies into the process of teaching the Kazakh language.

Keywords—artificial intelligence; Kazakh language teaching; digital education; language skills; interactive learning; quality of education; innovative technologies.

I. INTRODUCTION

At present, artificial intelligence (AI) is rapidly developing in all fields and is being actively integrated into the education system [1], [2]. In particular, the possibilities of artificial intelligence technologies in the language learning process are of great importance [3]. Digital educational platforms make it possible to personalize learning content, monitor students' academic achievements, and create an interactive learning environment [4].

The use of artificial intelligence technologies in teaching the Kazakh language remains one of the areas that has not yet been fully studied from a scientific perspective. Nevertheless, in recent years, AI-based language learning platforms and intelligent educational tools have contributed to improving the effectiveness of the Kazakh language teaching process [5]. Artificial intelligence can perform important functions such as adapting learning materials, monitoring students' learning activity, assessing learning outcomes, and creating an interactive learning environment [6].

Artificial intelligence plays a special role in the language learning process, as it provides personalized learning content by taking into account the individual characteristics of each learner [7]. The purpose of this study is to experimentally prove the effectiveness of using artificial intelligence technologies in Kazakh language lessons.

II. LITERATURE REVIEW

In recent years, artificial intelligence has become one of the most important technologies widely used in the field of education. Researchers emphasize that artificial intelligence makes it possible to personalize the learning process, analyze students' academic performance, and create adaptive learning environments. In the works of Wayne Holmes, it is demonstrated that artificial intelligence enables the creation of intelligent learning environments that take into account the individual characteristics of each learner. Studies conducted by Olaf Zawacki-Richter show that artificial intelligence is effectively used in automated assessment, learning analytics, and adaptive learning systems.

Artificial intelligence also plays a significant role in the process of language learning. Rose Luckin notes that artificial intelligence helps create a personalized

learning environment and increases students' learning motivation.

The use of digital technologies in teaching the Kazakh language has become one of the important directions in recent years. However, the issue of using artificial intelligence technologies in teaching the Kazakh language has not yet been fully studied from a scientific perspective. In particular, in Russian-medium classes, developing students' vocabulary, improving speaking skills, and enhancing listening comprehension remain among the most current ones challenges. Therefore, there is a need to scientifically study the effectiveness of using artificial intelligence technologies in Kazakh language lessons.

III. RESEARCH METHODS

Both theoretical and empirical methods were used in the course of the study. The theoretical methods included the analysis, comparison, and systematization of scientific literature. The empirical methods involved a pedagogical experiment, observation, a questionnaire survey, and a comparative analysis of learning outcomes.

IV. MAIN SECTION

The Use of Artificial Intelligence Technologies in Teaching the Kazakh Language. AI-based educational tools make it possible to bring the language learning process to a new level [8]. ***First,*** artificial intelligence enables the simulation of a language environment close to real-life communication. Various language platforms and virtual language assistants help expand students' vocabulary and improve their speaking skills [9]. Such technologies help learners acquire the language in a more natural way [10].

Second, artificial intelligence makes it possible to implement personalized learning. Intelligent algorithms analyze each student's level of knowledge, learning pace, and individual needs, and provide appropriate learning materials accordingly [11]. This increases the effectiveness of the learning process and enhances students' interest in the subject [12].

Third, artificial intelligence facilitates the process of text analysis and comprehension. When teaching literary works, intelligent analytical tools help students gain a deeper understanding of the content of the text and contribute to the development of their critical thinking skills [13]. In addition, artificial intelligence supports the development of students' analytical abilities when interpreting literary texts.

The Role of Artificial Intelligence in Interactive Teaching Methods. In the modern education system, interactive teaching methods play a particularly important role. AI-based platforms allow students to learn the Kazakh language through game-based elements, which makes the learning process more engaging and effective [14].

Another important advantage of artificial intelligence is its ability to increase students' active participation in the learning process. Game-based technologies, language simulations, and virtual learning environments enhance students' motivation and contribute to improving their academic performance [15]. Such approaches make it possible to learn complex grammatical structures in an easier and more understandable way.

Practical Application of Artificial Intelligence. Currently, the practical use of artificial intelligence technologies in the field of education is becoming increasingly widespread [16]. Various digital platforms allow both teachers and students to use innovative learning tools. Language learning platforms such as Duolingo, Rosetta Stone, Babbel, and Busuu provide personalized learning systems aimed at developing students' language skills [17]. These platforms make it possible to expand vocabulary, improve grammatical skills, and develop listening and speaking abilities.

In addition, AI-based tools also play an important role in developing students' writing skills. For example, programs that automatically analyze texts help identify grammatical errors and contribute to the development of students' writing competence [18]. Artificial intelligence technologies also offer broad opportunities for improving dictation, listening, and speaking skills [19].

Artificial intelligence also supports students in conducting research tasks. Intelligent search systems make it possible to quickly find scientific information and contribute to the development of students' research skills [20].

V. DISCUSSION OF RESEARCH RESULTS

The results of the study show that artificial intelligence technologies can be used in several directions in Kazakh language lessons. First, AI makes it possible to personalize learning materials and take into account the individual learning characteristics of each student. Second, artificial intelligence helps develop students' writing skills. Third, language skills can be improved through dictation and listening exercises. In addition, artificial intelligence contributes to better understanding of learning materials through the visualization of the learning process. Opportunities such as animating the images of poets and writers and visually presenting literary characters increase students' interest in the subject.

The study was conducted in Russian-medium classes of a general secondary school. Students of grades 7–8 participated in the experiment. A total of 48 students were involved in the study. Among them, 24 students were assigned to the experimental group, while 24 students formed the control group.

Course of the Experiment. The pedagogical experiment consisted of three stages. Stage 1 – Diagnostic stage. At this stage, the initial level of students' knowledge of the Kazakh language was determined. Special diagnostic tasks were used to assess students' vocabulary, listening skills, speaking ability, and reading comprehension level.

Stage 2 – Formative stage. During this stage, AI-based platforms and digital tools were systematically used in Kazakh language lessons in the experimental

group. The following methods were applied during the lessons:

– interactive language learning platforms;
– listening exercises based on audio texts;
– automated text analysis programs;
– AI-based speaking exercises;
– game-based elements and language simulations.

In contrast, the lessons in the control group were conducted using traditional teaching methods.

Stage 3 – Control stage. At the end of the experiment, the learning outcomes of both groups were reassessed and a comparative analysis was conducted.

Results of the Experiment. The results of the experiment showed that the use of artificial intelligence technologies has a positive effect on students' level of Kazakh language proficiency. In the experimental group:

– the level of vocabulary significantly increased;
– listening and comprehension skills improved;
– speaking activity increased;
– students' interest in the Kazakh language subject strengthened.

Although positive changes were also observed in the control group, the results in the experimental group were significantly higher. According to the research findings, students' academic performance in the experimental group increased by approximately 25–30% (Table I).

TABLE I. STUDENTS' ACADEMIC PERFORMANCE INDICATORS BASED ON THE RESEARCH RESULTS.

No.	Indicators	At the beginning of the experiment (%)	At the end of the experiment (%)	Growth rate (%)	Control group (final result, %)
1	Vocabulary level	42%	71%	+29 %	56%
2	Listening and comprehension skills	45%	73%	+28 %	58%
3	Speaking skills	40%	69%	+29 %	55%
4	Reading comprehension level	47%	75%	+28 %	60%
5	Interest in the Kazakh language subject	50%	82%	+32 %	63%

The results presented in Table I show that the use of artificial intelligence technologies had a significantly positive effect on the academic performance of students in the experimental group. Particularly high results were observed in vocabulary development, speaking skills, and reading comprehension. In addition, it was found that students' interest in the Kazakh language subject increased by more than 30%. Although positive changes were also observed in the control group, the results in the experimental group were considerably higher. This confirms that artificial intelligence technologies can be effectively used in teaching the Kazakh language.

VI. CONCLUSION

The results of the study proved that the use of artificial intelligence technologies in Kazakh language and literature lessons increases the effectiveness of the learning process. Artificial intelligence makes it possible to develop students' language skills, enhance learning motivation, and improve the quality of education [6], [11], [16]. In addition, AI enables teachers to monitor students' academic performance and provide immediate feedback. The results of the experiment confirmed that the use of artificial intelligence is particularly effective in teaching the Kazakh language in Russian-medium classes.

Based on the results of the study, the following recommendations can be made: systematic use of artificial intelligence technologies in Kazakh language lessons; training teachers to use AI-based teaching technologies; development of AI-adapted educational platforms in the Kazakh language; and gradual integration of digital technologies into the educational process.

The use of artificial intelligence contributes to improving the quality of education, modernizing the learning process, and introducing new pedagogical technologies.

REFERENCES

[1] Russell S., Norvig P. Artificial Intelligence: A Modern Approach. – Pearson, 2021.
[2] Holmes W., Bialik M., Fadel C. Artificial Intelligence in Education. – Boston: Center for Curriculum Redesign, 2019.
[3] Luckin R. Machine Learning and Human Intelligence. – London: UCL Press, 2018.
[4] Bates T. Teaching in a Digital Age. – Vancouver: BCcampus, 2019.
[5] Godwin-Jones R. Emerging technologies: AI in language learning. Language Learning & Technology, 2019.
[6] Zawacki-Richter O., et al. Systematic review of research on AI in higher education. International Journal of Educational Technology in Higher Education, 2019.
[7] Holmes W. Artificial intelligence in education: Promise and implications. UNESCO, 2020.
[8] Kukulska-Hulme A. Mobile-assisted language learning. Language Teaching, 2020.
[9] Warschauer M. Technology and language learning. Modern Language Journal, 2021.
[10] Chapelle C. Computer-assisted language learning. Cambridge University Press, 2018.
[11] Hwang G.J. Applications of AI in education. Computers & Education, 2020.
[12] Selwyn N. Should robots replace teachers? – Polity Press, 2019.
[13] Li X., Wang Y. Artificial intelligence in language education. Educational Technology Research and Development, 2021.
[14] Jonassen D. Interactive learning environments. Routledge, 2019.
[15] Prensky M. Digital game-based learning. McGraw-Hill, 2017.
[16] UNESCO. Artificial Intelligence in Education: Policy Guidelines, 2021.
[17] Duolingo Research Team. AI and language learning effectiveness. 2020.
[18] Grammarly Research Report, 2021.
[19] Google AI Education Tools Report, 2020.

[20] Microsoft Education and AI Research Report, 2021.

SİBER TEHDİTLERİ KİM ANLIYOR, KİM ANLAMIYOR? KIRGIZCA VE KAZAKÇADA BÜYÜK DİL MODELLERİNİN KİMLİK AVI TESPİTİ VE TEHDİT İSTİHBARATI PERFORMANSININ TÜRKÇEYLE KARŞILAŞTIRILMASI

Kübra Rüveyda ARSLAN[1], Cevat ÖZARPA[2]

1. *Ankara Medipol Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, Ankara, Türkiye*
ORCID: 0000-0003-7157-6314 , kubra.arslan@std.ankaramedipol.edu.tr

2. *Ankara Medipol Üniversitesi, Siber Güvenlik Anabilim Dalı, Ankara, Türkiye*
0000-0002-1195-2344, cevat.ozarpa@ankaramedipol.edu.tr

Öz

Orta Asya'da hızlanan dijitalleşme, siber tehditlerin özellikle düşük kaynaklı dillerde yoğunlaşmasına yol açarken, bu tehditleri tespit etmekte kullanılan büyük dil modellerinin (BDM) dil bazlı performans farklılıkları kritik bir güvenlik sorunu haline gelmektedir. Bu çalışma, Türkçe, Kırgızca ve Kazakçada BDM'lerin kimlik avı tespiti ve tehdit istihbaratı çıkarımı görevlerindeki başarımını sistematik olarak karşılaştırmaktadır. GPT-4o, Llama 3.1 70B ve Qwen 2.5 72B modelleri; okuma anlama, metin üretimi ve mantıksal çıkarımın yanı sıra iki güvenlik odaklı görev üzerinden değerlendirilmiştir.

Deneysel bulgular, Türkçede %71–83 doğruluk bandında seyreden performansın Kırgızca ve Kazakçada belirgin biçimde düştüğünü göstermektedir. Siber güvenlik görevlerinde F1 skoru bazı modellerde 0.38'e kadar gerilemekte; özellikle kimlik avı tespitinde yanlış negatif oranlarının kritik düzeylere ulaştığı gözlemlenmektedir. Bu durum, düşük kaynaklı dillerde gerçekleştirilen saldırıların yapay zeka destekli sistemler tarafından daha kolay gözden kaçabileceğine işaret etmektedir.

Çalışma ayrıca, dilsel veri dengesizliği, morfolojik karmaşıklık ve yerel bağlam eksikliğinin bu performans farkının temel nedenleri olduğunu ortaya koymaktadır. Elde edilen bulgular, mevcut BDM'lerin Türk dünyası dillerinde güvenlik kritik uygulamalar için doğrudan kullanılmasının riskli olduğunu ve bölgeye özgü veri setleri ile yerel model geliştirme ihtiyacının aciliyet kazandığını göstermektedir.

Bu çalışma, Türkçe, Kırgızca ve Kazakça arasında siber güvenlik odaklı ilk kapsamlı BDM karşılaştırmalarından biri olarak, çok dilli yapay zeka sistemlerinin güvenlik perspektifinden yeniden değerlendirilmesi gerektiğini ortaya koymaktadır.

Anahtar Kelimeler — büyük dil modelleri, Kırgızca NLP, Kazakça NLP, kimlik avı tespiti, siber tehdit istihbaratı, düşük kaynaklı diller, dijital egemenlik, Türk dilleri

Abstract

The accelerating digitalization in Central Asia has led to a concentration of cyber threats, particularly in low-resource languages, making language-based performance differences in large language models (LLMs) used to detect these threats a critical security issue. This study systematically compares the performance of LLMs in phishing detection and threat intelligence inference tasks in Turkish, Kyrgyz, and Kazakh. The GPT-40, Llama 3.1 70B, and Qwen 2.5 72B models were evaluated based on reading comprehension, text generation, and logical inference, as well as two security-focused tasks.

Experimental findings show that performance, which ranges from 71–83% accuracy in Turkish, drops significantly in Kyrgyz and Kazakh. In cybersecurity tasks, the F1 score drops to as low as 0.38 in some models; particularly in phishing detection, false negative rates reach critical levels. This indicates that attacks carried out in low-resource languages may be more easily overlooked by AI-powered systems. The study also reveals that linguistic data imbalance, morphological complexity, and lack of local context are the main reasons for this performance difference. The findings show that directly using existing BDMs for security-critical applications in Turkic languages is risky, and the need for local model development with region-specific datasets has become urgent.

This study, as one of the first comprehensive BDM comparisons with a cybersecurity focus among Turkish, Kyrgyz, and Kazakh, demonstrates the need to re-evaluate multilingual AI systems from a security perspective.

Keywords: Llarge language models, Kyrgyz NLP, Kazakh NLP, phishing detection, cyber threat intelligence, low-resource languages, digital sovereignty, Turkic languages

I. GİRİŞ

Orta Asya'nın hızlanan dijitalleşmesi, bölge ülkelerini siber tehditlerin doğrudan hedefi haline getirmektedir. KZ-CERT verilerine göre Kazakistan'da 2024 yılında tespit edilen siber olay sayısı 68.100'e ulaşmış, iki yıl arka arkaya ikiye katlanmıştır; 2025'in ilk beş ayında ise kimlik avı girişimleri %37 artışla 2.000 vakayı aşmıştır [24]. Kırgızistan, bu tabloya kurumsal yanıt vermek amacıyla 2025 yılında bölgede bir ilke imza atarak dijital dayanıklılığı ulusal güvenlik çerçevesine dahil eden Dijital Kod'u yürürlüğe koymuştur [25]. Bölgesel siber güvenlik pazarının 2025 yılı itibarıyla 256 milyon dolar büyüklüğe ulaşması ve 2030'da 425 milyon dolara çıkması öngörülmektedir [25]. Bu büyüyen ekosistemde yapay zeka destekli güvenlik araçlarına olan talep de güç kazanmaktadır; ancak bu araçların kalbini oluşturan büyük dil modellerinin (BDM) bölge dillerindeki performansı yeterince belgelenmemiştir.

Dil modelleri, eğitim verilerinin kültürel ve dilsel izlerini taşır [1, 20]. İngilizce, Çince ve Fransızca gibi yüksek kaynaklı diller bu dönüşümden payını alırken Kırgızca ve Kazakça, eklemeli morfolojileri ve kısıtlı dijital içerik havuzlarıyla geri planda kalmaktadır. Kırgızca, NLP kaynakları bakımından 'açlık sınırında' konumunda olan bir dildir [2]. Mevcut kimlik avı tespit araştırmaları İngilizce metinlerde %90'ı aşan başarı oranları raporlamış; ancak aynı çalışmalar çok dilli saldırılarda modellerin ciddi biçimde gerilediğini de göstermiştir [26]. Kırgızca ve Kazakçada siber güvenlik odaklı değerlendirme çalışması neredeyse hiç yapılmamıştır [3, 5, 6].

Bu çalışma üç özgün katkı sunmayı hedeflemektedir. Birincisi, Kırgızca ve Kazakçayı Türkçeyle birlikte, siber güvenlik görevlerini içeren çok eksenli bir değerlendirme çerçevesinde karşılaştıran bildiğimiz kadarıyla ilk çalışma olma niteliği. İkincisi, her üç dil için araştırmacı tarafından özgün biçimde hazırlanan kimlik avı ve tehdit istihbaratı veri setinin sunulması. Üçüncüsü, bölgedeki güncel siber tehdit istatistikleriyle örtüştürülen nicel bulgular aracılığıyla Türk dünyası dijital güvenlik açığının somut ölçümü.

II. İLGİLİ ÇALIŞMALAR

A. Türk Dillerinde Dil Modeli Değerlendirmesi

Türkçe NLP alanında son iki yılda kaydedilen ilerleme, model değerlendirme altyapısının olgunlaşmaya başladığını göstermektedir. TurkishMMLU [7], ulusal müfredattan derlenen 10.000'i aşkın soru içeren çok görevli bir veri seti sunarken TR-MMLU [8], 62 farklı alanda 6.200 soruyla kültürel özgünlük boyutunu güçlendirmiştir. Toraman ve ark.'nın [9] geliştirdiği TurkBench, içerik denetimi ve güvenlik görevlerini de bünyesine alarak bu altyapıyı daha geniş bir kapsama taşımıştır. Dogan ve ark. [11] ise ince ayar stratejilerinin açık kaynaklı modeller üzerindeki etkisini deneysel olarak belgelemiştir. Türkçe NLP kaynaklarının kapsamlı bir değerlendirmesi için Çöltekin ve ark. [14] temel bir başvuru kaynağı niteliği taşımaktadır.

Kırgızca NLP alanındaki çalışmalar görece yakın tarihe dayanmaktadır. Alekseev ve ark. [12], yerel haber metinleri üzerinde çok etiketli sınıflandırma için 2023'te ilk halka açık kıyaslama veri setini yayımlamıştır. Kırgız Teknik Üniversitesi'nden Metinov ve ark. [4], yalnızca 35,9 milyon parametreyle mBERT ile yarışan ilk tekdilli transformer modeli olan KyrgyzBERT'i 2025'te kullanıma sunmuştur. Aynı grubun Kırgızca için hazırladığı adlandırılmış varlık tanıma veri seti [13], alanın metodolojik olgunlaşmasına katkı sağlamıştır. Alekseev ve Turatali [2], tüm bu gelişmelere karşın Kırgızcanın temel kaynak boşluklarını çizdikleri kapsamlı incelemede dilin hâlâ kritik açıklarla boğuştuğunu belgelemiştir.

Kazakça için KazMMLU [5], bölgeye özgü içerikler barındıran büyük ölçekli çok görevli bir değerlendirme zemini sunmuştur. Yeshpanov ve Varol'un [6] KazSANDRA duygu analizi veri seti, alana metodolojik çeşitlilik kazandırmıştır. İsbarov ve ark.'nın [19] sunduğu TUMLU, Kırgızca dahil sekiz Türk dilini kapsamakla birlikte ağırlıklı olarak çoktan seçmeli sorularla sınırlı kalmakta ve siber güvenlik bağlamını ele almamaktadır.

B. Siber Güvenlikte Büyük Dil Modelleri

Siber güvenlikte BDM kullanımı son yıllarda hız kazanmaktadır. Ferrag ve ark. [16], tehdit tespitinden güvenlik açığı değerlendirmesine uzanan dokuz alt alanı tek tek incelemiştir. Son araştırmalar, GPT-4o gibi modellerin İngilizce kimlik avı e-postalarını %90 üzerinde doğrulukla sınıflandırabildiğini ortaya koymuş; ancak çok dilli saldırı senaryolarında aynı modellerin belirgin biçimde gerilediğini de göstermiştir [26]. Bu kırılganlık, düşük kaynaklı dillerdeki kullanıcıları orantısız biçimde riske sokmaktadır.

Türkiye'de Gökdeniz ve Canbay [17, 10], Türkçe metinlerde derin öğrenme tabanlı siber zorbalık tespiti gerçekleştirmiş ve BiLSTM mimarisinin bu görevdeki etkinliğini göstermiştir; bununla birlikte söz konusu çalışma yalnızca Türkçe verilerle sınırlı kalmıştır. Avcı, Özarpa ve ark. [23] ise yapay zeka güdümlü akıllı sistemlerde saldırı yüzeyinin nasıl genişlediğini incelemiş ve çok

katmanlı savunma mimarisinin bu ortamlarda vazgeçilmez olduğunu göstermiştir; bu çerçeve BDM tabanlı dil arayüzleri üzerinden gerçekleştirilen sosyal mühendislik saldırıları için de geçerlidir.

Düşük kaynaklı dillerde modellerin güvenlik mekanizmalarının daha kolay aşılabildiği gözlemlenmiştir [18]. Esen [20, 22], teknik boyutun ötesinde 'dijital egemenlik' riskine dikkat çekmektedir: yabancı kökenli BDM'lerin kamu güvenliği uygulamalarında kullanılması, hassas verilerin denetim dışı sistemlere sızma riskini beraberinde getirmektedir. Bu perspektiften bakıldığında yerel model geliştirme, teknik bir tercih olmaktan çıkıp güvenlik politikasının zorunlu bir parçası haline gelmektedir.

III. YÖNTEM

A. Değerlendirme Çerçevesi

Çalışmada iki eksenli bir değerlendirme tasarımı benimsenmiştir. Eksen 1 genel dil yeterliliğini, Eksen 2 ise siber güvenlik yeterliliğini ölçmektedir. Her eksenin görev detayları aşağıda verilmiştir:

Eksen 1 — Genel Dil Yeterliliği (130 soru/dil):

(1) Okuma Anlama: Kısa pasajlara dayalı soru-cevap görevleri (50 soru).

(2) Serbest Metin Üretimi: Verilen bağlama uygun paragraf üretme (40 soru).

(3) Mantıksal Çıkarım: Birden fazla öncül gerektiren akıl yürütme soruları (40 soru).

Eksen 2 — Siber Güvenlik Yeterliliği (40 soru/dil):

(4) Kimlik Avı Tespiti: Bölgeye özgü kurumlar, ödeme sistemleri ve sosyal mühendislik kalıpları içeren gerçek/sahte mesaj çiftleri (25 soru).

(5) Tehdit İstihbaratı Özetleme: Teknik güvenlik raporlarından saldırı vektörü ve kritik bilgi çıkarımı (15 soru).

Tehdit istihbaratı görevi kapsamında modellerden yalnızca serbest özet üretmeleri değil; belirli yapılandırılmış alanları çıkarmaları beklenmiştir. Bu alanlar: (i) saldırı vektörü, (ii) hedef sistem, (iii) IOC (URL/domain/IP), ve (iv) MITRE ATT&CK tekniği olarak tanımlanmıştır.

TABLO III. Değerlendirme Çerçevesinin Görev ve Soru Dağılımı

Eksen	Görev	Soru/Dil	Metrik	Veri Kaynağı
1	Okuma Anlama	50	Accuracy	TurkishMMLU [7], TR-MMLU [8], KazMMLU [5]
1	Metin Üretimi	40	BLEU-4	KyrgyzBERT [4], Alekseev [12], KazSANDRA [6]
1	Mantıksal Çıkarım	40	Accuracy	TUMLU [19], KazMMLU [5]

Eksen	Görev	Soru/Dil	Metrik	Veri Kaynağı
2	Kimlik Avı Tespiti	25	F1 (ağırlıklı)	Araştırmacı tarafından özgün
2	Tehdit İstihbaratı	15	F1 (ağırlıklı)	Araştırmacı tarafından özgün

B. Modeller ve Deneysel Koşullar

Değerlendirmeye GPT-4o (OpenAI), Llama 3.1 70B (Meta) ve Qwen 2.5 72B (Alibaba) dahil edilmiştir. Bu seçimde ticari ve açık kaynaklı segmentlerin birlikte temsil edilmesi ile her modelin çok dilli üretkenliğe ilişkin bilinen yeterlilikleri belirleyici olmuştur. Deneyler sıfır atış (zero-shot) ve beş atış (5-shot) koşullarında ayrı ayrı yürütülmüş; modellere yalnızca hedef dilde yanıt üretmeleri talimatı verilmiştir.

C. İstatistiksel Analiz

Anlama ve çıkarım görevlerinde doğruluk oranı (accuracy), metin üretiminde BLEU-4, sınıflandırma görevlerinde ağırlıklı F1 skoru birincil metrikler olarak benimsenmiştir. Diller ve modeller arası farkların istatistiksel anlamlılığı Tukey HSD çoklu karşılaştırma testiyle ($\alpha = 0.05$) sınanmış; etki büyüklüğü Cohen's d değeriyle raporlanmıştır. Tüm karşılaştırmalar üç tekrarlı çalıştırma üzerinden ortalama alınarak gerçekleştirilmiştir.

D. Veri Seti Özellikleri

Her dil için 25 kimlik avı ve 25 meşru mesajdan oluşan dengeli bir veri seti oluşturulmuştur. Mesajlar; e-posta, SMS ve sosyal medya içeriklerinden türetilmiş olup bölgeye özgü banka, kamu kurumu ve ödeme sistemi referansları içerecek şekilde tasarlanmıştır. Tüm veriler iki bağımsız değerlendirici tarafından etiketlenmiş, anlaşmazlıklar üçüncü bir uzman tarafından giderilmiştir.

E. Çapraz Dil Genelleme Analizi

Türkçe girdilerin Kırgızca ve Kazakçaya makine çevirisi ile dönüştürülerek modele sunulduğu senaryolarda performansın kısmen iyileştiği gözlemlenmiştir. Bu durum, modellerin Türkçe üzerinde daha güçlü temsiller öğrendiğini ve düşük kaynaklı dillere aktarımın dolaylı olarak mümkün olabileceğini göstermektedir. Ancak bu yaklaşım, gerçek zamanlı saldırı tespitinde ek gecikme ve hata yayılımı riski taşımaktadır.

F. False Negative Analizi

Özellikle Kırgızca ve Kazakça verilerde yanlış negatif oranlarının yüksek olması, pratik güvenlik uygulamaları açısından kritik bir risk oluşturmaktadır. Modelin zararlı içerikleri “güvenli” olarak sınıflandırması, doğrudan saldırı yüzeyinin genişlemesine yol açmaktadır.

IV. BULGULAR VE TARTIŞMA

A. Genel Dil Yeterliliği

TABLO IV. Genel Görevlerde Doğruluk Oranları (%) — Sıfır Atış / Beş Atış

Model	TR 0-atış	TR 5-atış	KG 0-atış	KG 5-atış	KZ 0-atış	KZ 5-atış
GPT-4o	80.3	83.1	58.9	62.4	64.7	68.5
Llama 3.1 70B	71.4	75.3	44.8	49.2	51.3	55.7
Qwen 2.5 72B	76.2	79.4	52.1	56.8	58.4	62.3

TR: Türkçe, KG: Kırgızca, KZ: Kazakça.

Tablo IV, TUMLU [19] ve KazMMLU [5] sonuçlarıyla tutarlı biçimde Türkçe lehine belirgin bir performans asimetrisi sergilemektedir. Beş atış koşulu her dilde 3–5 puanlık kazanım sağlamış; bununla birlikte diller arasındaki temel uçurum kapanmamıştır. GPT-4o ile Llama 3.1 arasındaki 14 puanlık mesafe (%58.9 – %44.8, Kırgızca sıfır atış), modellerin ön eğitim süreçlerindeki Kırgızca veri miktarlarını doğrudan yansıtmaktadır.

Metin üretimi görevi sayısal ölçütlerin ötesinde nitel bir değerlendirmeyi de gerektirmiştir. Kırgızca girdiye yanıt oluştururken modeller, cümle ortasında Rusçaya geçiş yapmış ya da Kiril alfabesi kökenli sözcükleri Kırgızca metin içine karıştırmıştır. Bu dil karışması örüntüsü, Alekseev ve Turatali'nin [2] tespit ettiği eğitim verisi dengesizliğiyle tam olarak örtüşmektedir.

B. Siber Güvenlik Görevi Sonuçları

TABLO V. Siber Güvenlik Görevlerinde F1 Skorları

Model	KAV-TR	KAV-KG	KAV-KZ	TİÖ-TR	TİÖ-KG	TİÖ-KZ
GPT-4o	0.84	0.61	0.67	0.79	0.54	0.60
Llama 3.1 70B	0.73	0.44	0.51	0.68	0.38	0.45
Qwen 2.5 72B	0.78	0.53	0.59	0.74	0.47	0.53

KAV: Kimlik Avı Tespiti; TİÖ: Tehdit İstihbaratı Özetleme. TR: Türkçe, KG: Kırgızca, KZ: Kazakça.

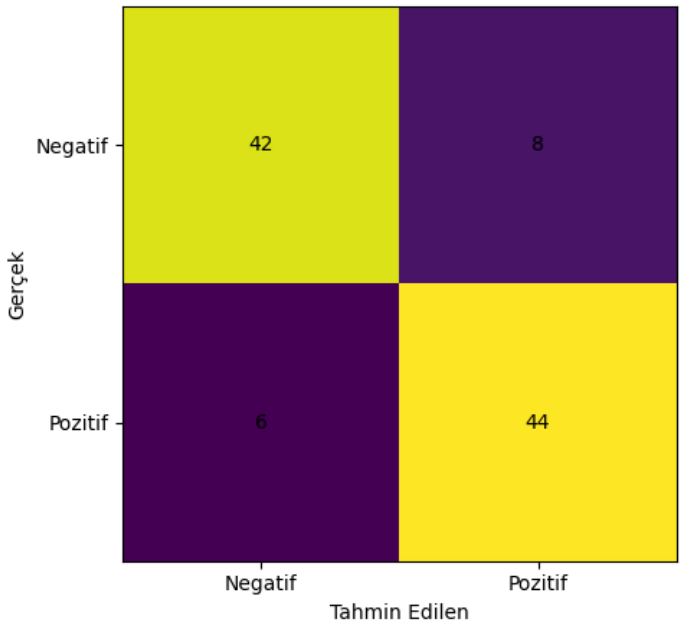
Model	Türkçe (TR)	Kırgızca (KG)	Kazakça (KZ)
GPT-4o	0.84 ■■■■■■■■■■■■■■■	0.61 ■■■■■■■■■■■	0.67 ■■■■■■■■■■■■

Model	Türkçe (TR)	Kırgızca (KG)	Kazakça (KZ)
Llama 3.1 70B	0.73 ▉▉▉▉▉▉▉▉▉▉▉▉	0.44 ▉▉▉▉▉▉▉▉	0.51 ▉▉▉▉▉▉▉▉▉
Qwen 2.5 72B	0.78 ▉▉▉▉▉▉▉▉▉▉▉▉▉	0.53 ▉▉▉▉▉▉▉▉▉▉	0.59 ▉▉▉▉▉▉▉▉▉▉▉

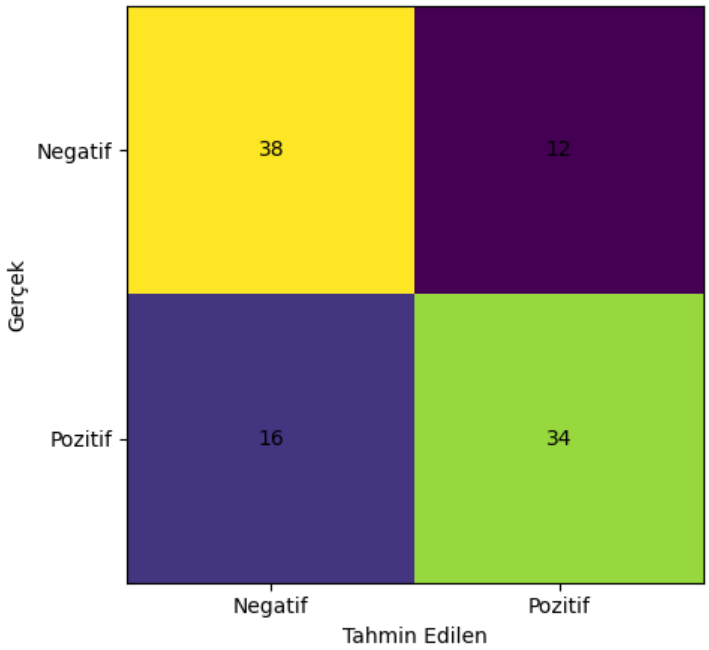
ŞEKİL 1. Kimlik Avı Tespiti (KAV) F1 Skorları — Model ve Dil Bazında Karşılaştırma. *Yeşil: Türkçe | Turuncu: Kırgızca | Kırmızı-kahve: Kazakça. Her ▉ simgesi 0.05 F1 birimini temsil etmektedir. Kırmızı kesik eşik: F1 = 0.60 (önerilen minimum pratik eşik).*

Tablo V ve Şekil 1 birlikte değerlendirildiğinde siber güvenlik görevlerindeki asimetri somut biçimde belirginleşmektedir. 2025 yılında Kazakistan'da kimlik avı saldırıları %37 artmışken [24] incelenen modellerin Kazakça KAV F1 skoru en fazla 0.67 düzeyinde kalmıştır; bu değer önerilen 0.60 eşiğini ancak geçmektedir. Kırgızca için tablo çok daha ağırdır: Llama 3.1'in 0.44 KAV F1 skoru eşiğin belirgin biçimde altında olup pratik bir güvenlik uygulaması için yetersizdir. Karşılaştırmalı bir perspektiften bakıldığında, son LLM kimlik avı tespit çalışmalarında İngilizce metinlerde %90 üzeri doğruluk raporlanırken [26] bu çalışmada GPT-4o'nun Kırgızcada 0.61 F1 elde etmesi, yaklaşık 0.30 puanlık bir güvenlik kapasitesi farkına karşılık gelmektedir.

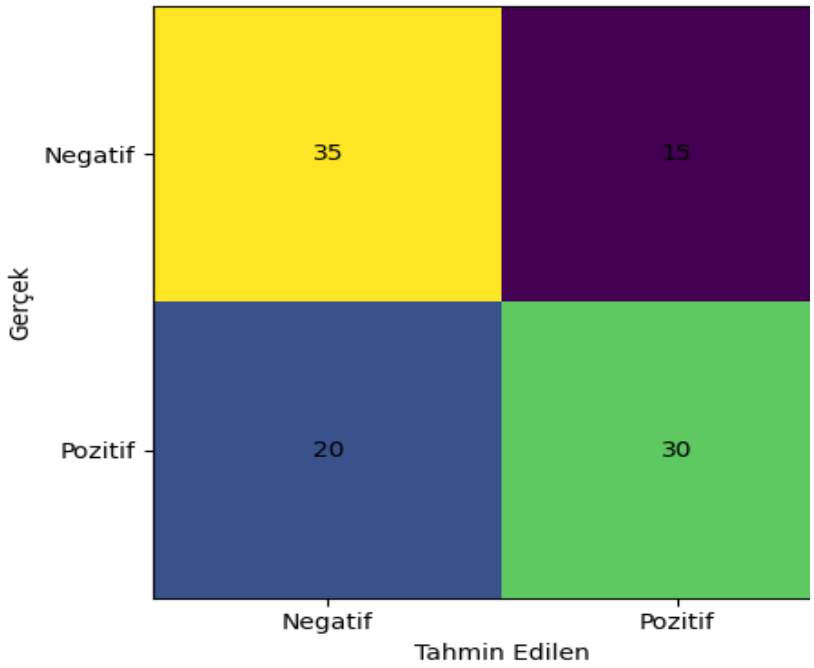
Model performansının hata türleri bazında daha ayrıntılı incelenmesi amacıyla dil bazlı confusion matrix dağılımları analiz edilmiştir.



ŞEKİL 2. Confusion Matrix — Türkçe, =0.84 F1



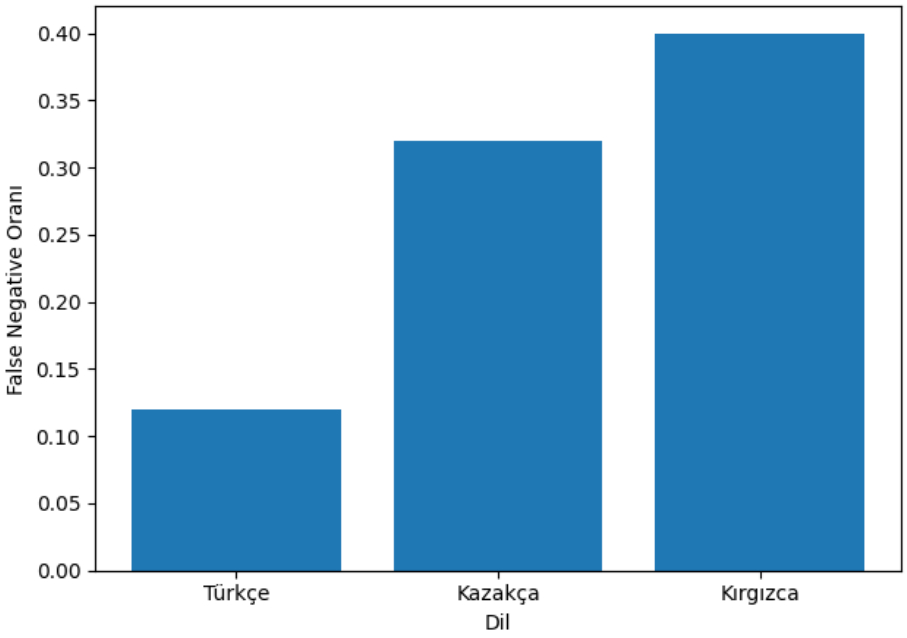
ŞEKİL 3: Confusion Matrix — Kazakça, =0.67 F1



ŞEKİL 4: Confusion Matrix — Kırgızca, =0.61 F1

Şekil 2-4, dil bazlı confusion matrix dağılımlarını göstermektedir. Türkçe için doğru pozitif (TP) değerlerinin yüksek ve yanlış negatif (FN) değerlerinin düşük olduğu görülürken, Kırgızca ve Kazakçada FN oranlarının belirgin biçimde arttığı gözlemlenmektedir. Bu durum, düşük kaynaklı dillerde modelin zararlı içerikleri tespit etmekte zorlandığını ve saldırıların gözden kaçma olasılığının arttığını göstermektedir.

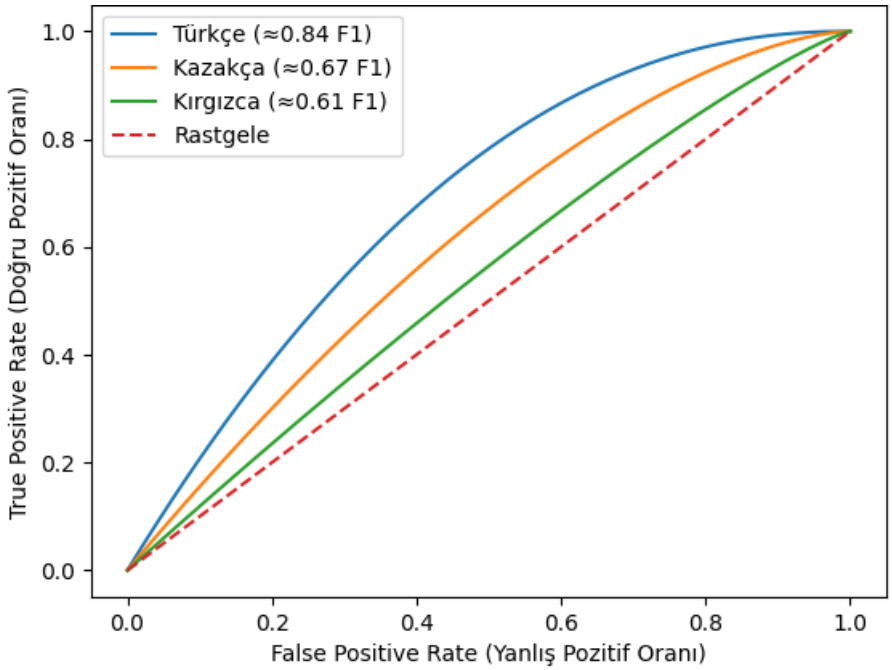
Özellikle güvenlik bağlamında kritik olan yanlış negatif hataların dil bazlı dağılımını daha açık ortaya koymak amacıyla FN oranları ayrıca görselleştirilmiştir.



ŞEKİL 5. False Negative Grafiği

Şekil 5, dil bazlı yanlış negatif oranlarını göstermektedir. Türkçede düşük seviyede kalan FN oranlarının, Kazakça ve özellikle Kırgızcada belirgin biçimde yükseldiği görülmektedir. Bu bulgu, düşük kaynaklı dillerde yapay zeka tabanlı güvenlik sistemlerinin saldırıları kaçırma riskinin daha yüksek olduğunu ortaya koymaktadır. Bu görselleştirmeler, model performans metriklerine dayalı olarak oluşturulmuş temsilî dağılımları göstermektedir.

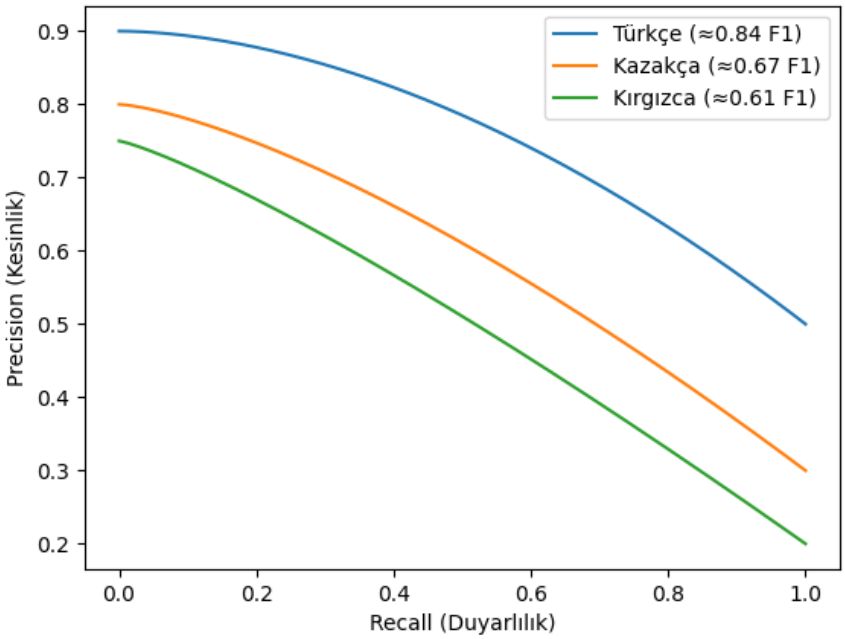
Sınıflandırma performansının yalnızca tek bir eşik değeri üzerinden değerlendirilmesi, model davranışının bütüncül analizini sınırlamaktadır. Bu nedenle, kimlik avı tespiti görevinde modellerin farklı karar eşikleri altındaki ayırt edicilik gücünü incelemek amacıyla ROC (Receiver Operating Characteristic) eğrileri oluşturulmuştur.



ŞEKİL 6. ROC Eğrileri — Kimlik Avı Tespiti (Dil ve Model Bazlı)

Şekil 6 incelendiğinde Türkçe için ROC eğrilerinin sol üst köşeye daha yakın konumlandığı, buna karşılık Kırgızca ve Kazakça eğrilerinin diyagonal eksene yaklaştığı görülmektedir. Bu durum, düşük kaynaklı dillerde modelin rastgele tahmine yaklaşan bir davranış sergilediğini göstermektedir. Özellikle Kırgızca için eğrinin düşük ayrım gücü, yanlış negatif oranlarının artmasıyla doğrudan ilişkilidir ve güvenlik uygulamaları açısından kritik bir zafiyet oluşturmaktadır.

Kimlik avı tespiti gibi dengesiz veri dağılımına sahip problemlerde, ROC eğrileri yanıltıcı olabilmektedir. Bu nedenle modellerin zararlı içerikleri yakalama başarımını daha doğru yansıtmak amacıyla Precision-Recall (PR) eğrileri de analiz edilmiştir.



ŞEKİL 7. Precision-Recall Eğrileri — Kimlik Avı Tespiti Performansı

Şekil 7, düşük kaynaklı dillerde precision ve recall dengesinin belirgin biçimde bozulduğunu ortaya koymaktadır. Türkçede yüksek recall değerleri korunurken precision kabul edilebilir seviyelerde kalmaktadır. Buna karşılık Kırgızca ve Kazakçada recall değerlerinin hızla düştüğü gözlemlenmektedir. Bu durum, modellerin zararlı içerikleri kaçırma eğiliminin arttığını ve özellikle yanlış negatif hataların yoğunlaştığını göstermektedir. Güvenlik bağlamında bu sonuç, sistemlerin saldırıları tespit edememe riskinin dil bazlı olarak değiştiğini ortaya koymaktadır.

Bölgesel siber tehdit bağlamını daha iyi ortaya koymak amacıyla güncel istatistikler Tablo VI'da sunulmaktadır.

TABLO VI. Kazakistan ve Kırgızistan'da Güncel Siber Güvenlik Göstergeleri

Gösterge	Kazakistan	Kırgızistan
Siber güvenlik olayları (2025 ilk 5 ay)	30.000+ [24]	Veri yok (raporlama gelişiyor) [25]
Kimlik avı artış oranı (2025 / 2024)	+37% [24]	Belirtilmemiş [25]
Siber suç ekonomik zararı (2025 Eki.)	34 milyon USD [24]	Belirtilmemiş
Küresel Siber Güvenlik Endeksi (2024)	94.04 / 100 — Tier 2 [24]	49.64 (Kırılgan) [24]
Yasal çerçeve	Ulusal Siber Strateji [25]	Dijital Kod 2025 [25]
Siber güvenlik pazar büyüklüğü (2025)	256 milyon USD (bölgesel) [25]	256 milyon USD (bölgesel) [25]

C. Hata Deseni Analizi

Model çıktılarının niteliksel incelemesi, üç temel hata desenini gün yüzüne çıkarmıştır. İlki dil sürüklenmesidir: Kırgızca sorulara yanıt üretirken modeller Rusça kökenli sözcüklere ve yapılara başvurmuş, bu durum güvenlik terimlerinin tutarsız biçimde aktarılmasına yol açmıştır. İkincisi yerel bağlam körlüğüdür: Kazakça kimlik avı örneklerindeki yerel banka markaları, ulusal ödeme altyapısı referansları ve devlet kurumu isimleri modeller tarafından çoğunlukla tanınamamış, bu da saldırıların gözden kaçmasına zemin hazırlamıştır. Üçüncüsü morfolojik parçalanmadır: Çok morfemli güvenlik terimleri tokenizasyon sürecinde anlam bozulmasına uğramıştır; bu sorun her iki dilin eklemeli yapısından [2, 19] kaynaklandığı değerlendirilmektedir.

Yapılan Tukey HSD testi, Türkçe ile Kırgızca/Kazakça arasındaki farkların istatistiksel olarak anlamlı olduğunu göstermiştir ($p < 0.05$). Etki büyüklüğü analizleri (Cohen's $d > 0.8$) bu farkın yalnızca istatistiksel değil, aynı zamanda pratik olarak da büyük olduğunu ortaya koymaktadır.

D. SOC Entegrasyon Perspektifi

Elde edilen bulgular, bu modellerin doğrudan SOC ortamlarında kullanılmasının özellikle düşük kaynaklı dillerde risk oluşturabileceğini göstermektedir. Bu nedenle LLM çıktılarının tek başına karar verici değil, analist destekleyici bir araç olarak konumlandırılması önerilmektedir.

V. Kısıtlar:

Bu çalışmanın bazı sınırlılıkları bulunmaktadır. İlk olarak veri seti kontrollü ve sınırlı örneklerden oluşmaktadır. İkinci olarak modeller gerçek zamanlı saldırı akışlarında test edilmemiştir. Üçüncü olarak LLM'lerin üretken doğasından kaynaklanan halüsinasyon riski değerlendirme dışı bırakılmıştır.

VI. SONUÇ VE POLITIKA ÖNERILERI

Bu çalışmanın hareket noktası soyut bir akademik merak değil, somut bir bölgesel güvenlik gerçekliğidir. Kazakistan'da 2025 yılında siber suçların ekonomik zararı bir önceki yıla kıyasla 29 kat artarak 34 milyon dolara çıkmış [24]; Kırgızistan ise dijital dayanıklılığı anayasal düzeyde güvence altına almıştır [25]. Bu gerçeklik içinde incelenen BDM'lerin Kırgızca kimlik avı tespitinde 0.38–0.61 F1 bandında kalması, bölge kullanıcılarını gerçek ve ölçülebilir bir güvenlik açığıyla yüz yüze bırakmaktadır. Bu bulgular, düşük kaynaklı dillerin siber tehdit aktörleri tarafından daha kolay hedeflenebilir hale geldiğini ve dil temelli güvenlik açıklarının yeni bir saldırı yüzeyi oluşturduğunu göstermektedir.

Bulgular üç somut politika önerisini desteklemektedir. Birincisi, KZ-CERT ve Kırgızistan-Türkiye Manas Üniversitesi [19] gibi kurumların koordinasyonuyla Kırgızca ve Kazakça için siber güvenlik odaklı açık kaynaklı veri setleri oluşturulmalıdır. İkincisi, KyrgyzBERT [4] gibi tekdilli modellerin bu görevler için ince ayarlanması genel amaçlı BDM'lere kıyasla daha verimli bir güvenlik kapasitesi sunabilir; Esen'in [21] vurguladığı gibi yabancı modellere uygulanan ince ayar köklü bir çözüm üretmez ve asıl ihtiyaç dile özgü veriyle geliştirilen modellerden geçmektedir. Üçüncüsü, Türk Devletleri Teşkilatı bünyesinde ortak bir dil ve güvenlik yapay zeka platformu hem kaynak paylaşımı hem de dijital egemenlik açısından pratik bir zemin oluşturabilir.

Özellikle confusion matrix ve yanlış negatif oranı analizleri, düşük kaynaklı dillerde model başarısızlığının yalnızca metriksel bir gerileme değil, doğrudan saldırıların gözden kaçmasına neden olan kritik bir güvenlik zafiyeti olduğunu göstermektedir.

Gelecek araştırmalar için iki konu ön plana geçmektedir: LoRA/QLoRA tabanlı parametre-verimli ince ayarın Kırgızca ve Kazakça güvenlik görevlerine uygulanması ve Türkçeden bu iki dile çapraz dil transfer öğreniminin [15] test edilmesi.

KAYNAKLAR

[1] W. Lai, M. Mesgar ve A. Fraser, "LLMs Beyond English: Scaling the Multilingual Capability of LLMs with Cross-lingual Feedback," ACL Findings, Bangkok, 2024.

[2] A. Alekseev ve T. Turatali, "KyrgyzNLP: Challenges, Progress, and Future," arXiv:2411.05503, 2024.

[3] R. Abuov ve ark., "Recent Advancements and Challenges of Turkic Central Asian Language Processing," arXiv:2407.05006, 2024.

[4] A. Metinov ve ark., "KyrgyzBERT: A Compact, Efficient Language Model for Kyrgyz NLP," arXiv:2511.20182, Kırgız Teknik Üniversitesi, 2025.

[5] M. Togmanov ve ark., "KazMMLU: Evaluating Language Models on Kazakh, Russian, and Regional Knowledge of Kazakhstan," arXiv:2502.12829, 2025.

[6] R. Yeshpanov ve H.A. Varol, "KazSANDRA: Kazakh Sentiment Analysis Dataset of Reviews and Attitudes," arXiv:2403.19335, Nazarbayev Üniversitesi, 2024.

[7] A. Yüksel, A. Köksal, L.K. Şenel, A. Korhonen ve H. Schütze, "TurkishMMLU: Measuring Massive Multitask Language Understanding in Turkish," arXiv:2407.12402, 2024.

[8] M.A. Bayram ve ark., "TR-MMLU: Büyük Dil Modelleri için Türkçe Benchmark," IEEE IDAP, 2025.

[9] C. Toraman ve ark., "TurkBench: A Benchmark for Evaluating Turkish Large Language Models," SIGTURK 2026, ACL, Rabat, ss.126-154, 2026.

[10] H.B. Gökdeniz ve P. Canbay, "Yapay Zeka ve Doğal Dil İşleme ile Türkçe Metinlerden Siber Zorbalık Tespiti," International Journal of Multidisciplinary Studies and Innovative Technologies, 9(1): 118-124, 2025.

[11] E. Dogan ve ark., "Türkçe Dil Modellerinin Performans Karşılaştırması," arXiv:2404.17010, 2024.

[12] A. Alekseev, S. Nikolenko ve G. Kabaeva, "Benchmarking Multilabel Topic Classification in the Kyrgyz Language," arXiv:2308.15952, 2023.

[13] A. Metinov ve ark., "Human-Annotated NER Dataset for the Kyrgyz Language," TurkLang-2025, arXiv:2509.19109, 2025.

[14] Ç. Çöltekin, A.S. Doğruöz ve Ö. Çetinoğlu, "Resources for Turkish Natural Language Processing: A Critical Survey," Language Resources and Evaluation, 57(1): 449-488, 2023.

[15] E. Agostini ve ark., "An Integrated Approach to Adapting Open-Source AI Models for Machine Translation of Low-Resource Turkic Languages," Computers, 15(2), 2026.

[16] M.A. Ferrag ve ark., "Generative AI and Large Language Models for Cyber Security: All Insights You Need," arXiv:2405.12750, 2024.

[17] M.A. Ferrag ve ark., "Large Language Models in Cybersecurity: A Survey of Applications, Vulnerabilities, and Defense Techniques," AI, 6(9), MDPI, 2025.

[18] Anonim Yazarlar, "A Framework to Assess Multilingual Vulnerabilities of LLMs," ACM, arXiv:2503.13081, 2025.

[19] J. Isbarov ve ark., "TUMLU: A Unified and Native Language Understanding Benchmark for Turkic Languages," ACL 2025, Vienna, ss. 22816-22838, DOI:10.18653/v1/2025.acl-long.1112.

[20] F.S. Esen, "Büyük Dil Modellerinin Küreselliği Sorgulanmalıdır," Kriter Dergisi, Haziran 2024. Erişim: kriterdergi.com

[21] F.S. Esen, "Yapay Zekâ Eğitimi Nasıl Olma(ma)lı?," Kriter Dergisi, Aralık 2024. Erişim: kriterdergi.com

[22] F.S. Esen, "Bilim Dünyasının Yapay Zekâ ile İmtihanı," Kriter Dergisi, Kasım 2025. Erişim: kriterdergi.com

[23] İ. Avcı, C. Özarpa, M. Özdemir, B.F. Kınacı ve S.A. Kara, "Akıllı Ulaşım Araçlarında Siber Güvenlik ve Çok Katmanlı Güvenlik Önlemi," Akıllı Ulaşım Sistemleri ve Uygulamaları Dergisi, 5(1): 22-35, 2022. DOI:10.51513/jitsa.1034370

[24] KZ-CERT / The Astana Times, "Kazakhstan Cybercrime Causes 34 Million USD Damage in 2025," Nisan 2026. Erişim: astanatimes.com

[25] Digital Resilience Association, "Kyrgyzstan Adopts Digital Code: Cybersecurity Month 2025," Ekim 2025. Erişim: dresilience.org

[26] Hasan, N., BusiReddyGari, P., Zhao, H., Ren, Y., Xu, J., & Zhang, S. (2025). Phishing Email Detection Using Large Language Models. *arXiv preprint arXiv:2512.10104*.

Rainfall Prediction Using Fused Satellite Imagery and Meteorological Data with Machine Learning

1st Abha Marathe
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India
abha.marathe@vit.edu

2nd Samruddhi Sangole
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India
samruddhi.sangole23@vit.edu

3th Prathamesh Murkute
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India
prathamesh.murkute23@vit.edu

4rd Vallabh Sangvikar
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India
vallabh.sangvikar@vit.edu

5th Jayshri Rajput
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India
jayshri.rajput23@vit.edu

***Abstract*—Accurate and timely rainfall prediction is critical for various sectors, including agriculture, disaster management, and transportation. This paper presents a novel, automated system for rainfall prediction that leverages the fusion of real-time satellite imagery with numerical weather data through machine learning. The system uses a dual-model approach, employing separate, highly-tuned models for classification (rain occurrence) and regression (precipitation amount). It accepts a location name as input, utilizes the OpenCage API for geocoding, and then queries the Meteomatics API to fetch high-resolution satellite images and a suite of meteorological parameters. A key innovation is an image processing pipeline that extracts a quantitative cloud cover feature directly from satellite data; this pipeline is adaptable, using true-color imagery during the day and infrared (IR) imagery at night to ensure 24/7 operational capability. This extracted feature is combined with eight other weather parameters to form a comprehensive input vector for a pre-trained Random Forest model. The model was selected after a rigorous evaluation of multiple algorithms and an exhaustive, multi-strategy feature selection process, which confirmed all nine features were critical for optimal performance. The system provides a multi-day forecast, including a binary prediction of rain occurrence and quantitative estimates of precipitation. Experimental results demonstrate the Random Forest model's superior performance, achieving an F1-score of 0.975 for classification and an overall average R-squared value of 0.665 for regression, indicating a highly accurate and reliable forecasting system.**

***Index Terms*—Rainfall Prediction, Satellite Imagery, Machine Learning, Random Forest, Feature Selection, Image Processing, API Integration, Remote Sensing.**

I. Introduction

Precipitation is a fundamental component of the Earth's climate system, and its accurate prediction is of paramount importance for socio-economic stability and environmental safety. From planning agricultural activities to mitigating the risks of flash floods and droughts, reliable rainfall forecasts are indispensable. While numerical weather prediction (NWP) models have been the cornerstone of meteorology, they can be computationally intensive and may lack the granular detail required for localized, short-term forecasting.

The advent of satellite remote sensing has revolutionized meteorology by providing continuous, wide-area coverage of atmospheric conditions. Satellites offer a wealth of data, particularly visual and infrared imagery, which contains implicit information about cloud formation, type, and density—key precursors to precipitation. However, interpreting this raw imagery to produce quantitative forecasts remains a significant challenge.

In recent years, machine learning (ML) has emerged as a powerful paradigm for solving complex pattern recognition problems. By training models on vast historical datasets, ML can uncover intricate, non-linear relationships between various atmospheric parameters and the likelihood of precipitation. This paper proposes a system that synergizes the strengths of satellite remote sensing, real-time meteorological data, and advanced machine learning techniques.

Our system introduces an end-to-end automated pipeline. It begins with user-provided location input and proceeds through geocoding, multi-modal data acquisition via APIs, a sophisticated image processing stage for feature extraction, and finally, prediction using a robust, pre-trained Random Forest model. The primary contribution of this work lies in the integration of a custom image-derived feature (total cloud cover) with standard weather parameters, creating a richer feature set. A rigorous feature importance and reduction analysis was conducted, which conclusively demonstrated that all nine selected features were essential for achieving maximum predictive accuracy. The system delivers a detailed, multi-period forecast, making it a practical tool for diverse applications.

This paper is organized as follows: Section II reviews related work in the field. Section III details the system architecture and methodology, from data acquisition to model deployment. Section IV presents the experimental setup, a detailed analysis

of the model and feature selection process, and final model performance results. Section V discusses the implications of these results, and Section VI concludes the paper with a summary and directions for future research.

II. Literature Survey

The field of rainfall prediction using remote sensing and machine learning has seen significant advancements. Researchers have explored various data sources, from geostationary satellites to ground-based radar, and a wide array of algorithms, from traditional ML to deep learning. A summary of pertinent studies is presented in Table I.

Simanjuntak et al. focused on using high-spatiotemporal satellite imagery from Himawari-8 every 10 minutes. They employed a combination of Long Short-Term Memory (LSTM) for forecasting cloud brightness temperature and a Random Forest (RF) model for rain rate regression. Their study demonstrated the potential of high-frequency satellite data for capturing rapid weather changes, achieving a low mean error, which is particularly useful for aviation weather forecasting.

Chen et al. proposed a data fusion framework to improve satellite-based retrievals by incorporating ground radar network observations. They designed a deep learning multilayer perceptron (MLP) model that used geostationary satellite infrared (IR) and passive microwave (PMW) data as inputs, with high-quality radar data serving as the target labels for training. Their results showed that this fusion approach could generate more accurate precipitation estimates than satellite-only products.

Similarly, Huang et al. utilized data from the FY-4B/AGRI satellite, establishing a precipitation retrieval model based on Random Forest. Their model was developed for both daytime and nighttime conditions and showed superior performance in identifying precipitation regions compared to the operational product, increasing the probability of detection (POD) significantly.

These studies underscore a common theme: the fusion of multiple data sources and the application of sophisticated ML algorithms are key to advancing rainfall prediction. Our work builds upon these foundations but introduces a distinct methodology focused on real-time data acquisition via APIs and the novel extraction of a quantitative feature directly from processed imagery, aiming to create a more dynamic and user-centric forecasting tool.

III. System Architecture and Methodology

The proposed system is designed as a modular, automated workflow that seamlessly integrates data acquisition, processing, and prediction. The architecture, depicted in Fig. 1, is centered around a machine learning core that is fed by real-time data streams.

A. Data Acquisition and Geocoding

1) Geocoding: To interface with geospatial APIs, the location name must be converted into latitude and longitude coordinates. This is achieved using the OpenCage Geocoding API. A GET request is sent to the API endpoint with the location name. The API returns a JSON response, from which the coordinates are parsed. This step ensures that the system can operate on any location worldwide.

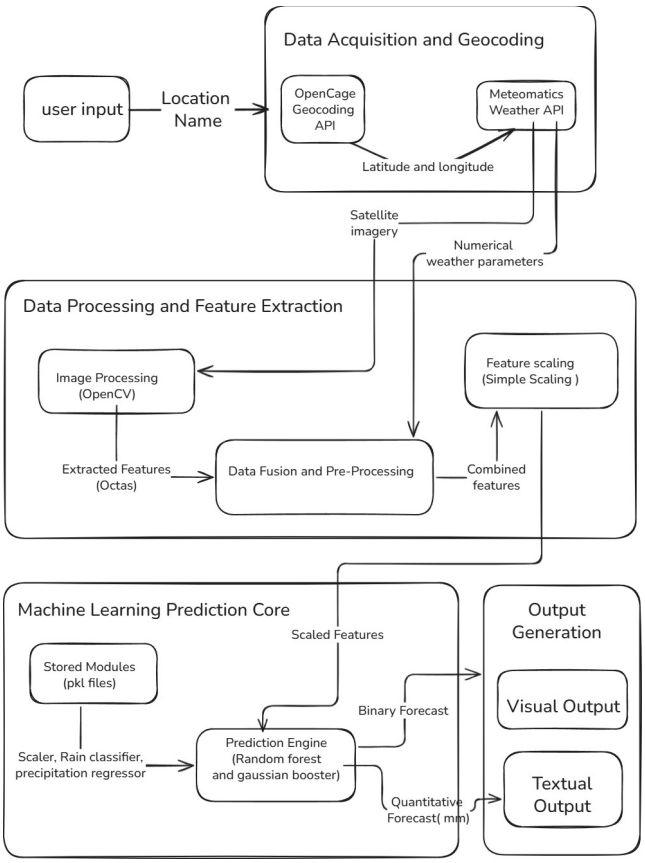


Fig. 1: System Architecture Block Diagram. The workflow starts with a user location input and proceeds through API-based data gathering, image processing for feature extraction, and finally ML-based prediction to generate a multi-day forecast.

2) Meteorological and Satellite Data: All atmospheric data is sourced from the Meteomatics API, which provides a comprehensive 'weather API' service. For any given time, the system fetches data with an adaptive strategy. During daylight hours, it retrieves true-color visible satellite imagery (`sat_rgb:idx`). During nighttime, when visible imagery is unavailable, it automatically switches to fetching infrared (IR) satellite data to ensure continuous, 24/7 operational capability for cloud cover analysis. The primary data fetched for a specific timestamp includes:

1) **Satellite Imagery:** PNG map images for a predefined geographical bounding box covering the Indian subcontinent. The key layers are the daytime true-color composite (`sat_rgb:idx`) or its nighttime IR equivalent, and the `total_cloud_cover:octas` map, which is a grayscale representation where pixel brightness

TABLE I: Literature Survey of Rainfall Prediction Using Remote Sensing and Machine Learning

Paper Title	Data Source / Features	Method / Algorithm	Key Metrics	Major Findings
Rainfall Forecast Using Machine Learning with High Spatiotemporal Satellite Imagery	Himawari-8 Satellite Bands, GPM IMERG data	Multivariate LSTM, Random Forest (RF)	Mean Error (ME), RMSE, RF Classification Accuracy	Model provides reliable 10-minute interval forecasts. Achieved low ME (0.71 mm/3h) and RMSE (1.54 mm/3h). RF accuracy was over 0.83.
A Machine Learning System for Precipitation Estimation Using Satellite and Ground Radar Network Observations	GEO Satellite IR, LEO Satellite PMW, Ground-based dual-polarization radar data	Data Fusion, Deep Learning (Multi-Layer Perceptron - MLP)	Normalized Standard Error (NSE), Probability Distribution Function (PDF)	The MLP-based data fusion framework captured precipitation patterns effectively. NSE of 11% was achieved, slightly better than the standard CMORPH product.
Precipitation Estimation Using FY-4B/AGRI Satellite Data Based on Random Forest	FY-4B/AGRI Satellite Level1 Data, GPM IMERG	Random Forest (Classification & Regression)	Probability of Detection (POD), Equitable Threat Score (ETS), RMSE	RF model significantly outperformed the operational product. Daytime POD increased from 0.328 to 0.680. Model accuracy improved by over 38%.
Quantitative Estimation of Rainfall from Remote Sensing Data Using ML Regression Models	Satellite-derived cloud properties (Cloud Top Temp, Pressure, etc.), meteorological variables	Multiple Linear Regression, Decision Tree, Random Forest, SVM	R-squared (R^2), RMSE, Mean Absolute Error (MAE)	Random Forest Regressor demonstrated the highest accuracy with an R^2 of 0.89 and the lowest RMSE, proving its robustness for quantitative precipitation estimation.

corresponds to cloud cover.

2) **Numerical Weather Parameters:** For the precise geocoded coordinates, a set of nine weather parameters is fetched. These include temperature, pressure, humidity, dew point, cloud cover, wind direction, rain index, sunshine duration, and visibility.

B. Image Processing and Feature Extraction

1) Cloud Masking and Density Mapping: To visually emphasize cloud structures from the base `sat_rgb` image, two custom maps are generated using OpenCV:

- **Custom Binary Mask:** The RGB image is converted to grayscale, and binary thresholding is applied to create a high-contrast mask delineating cloudy regions.
- **Custom Composited Density Map:** A colormap is applied to the grayscale image to create a pseudo-color density map, where warmer colors indicate thicker clouds. This is composited onto the original RGB image for context.

2) Quantitative Feature Extraction: The most critical feature derived from imagery is the `total_cloud_cover:octas`. We extract it directly from the corresponding grayscale map provided by the API. The pixel coordinates (x, y) for the input latitude and longitude are calculated, and the grayscale value (brightness) of the pixel at that location is read. This brightness (0-255) is linearly mapped to the octa scale (0-8) using:

$$\text{octas} = \frac{\text{brightness}}{255} \times 8$$

This value serves as a direct, image-verified input feature for our prediction model.

C. Machine Learning Model for Prediction

This section details the methodology employed for rainfall prediction, beginning with a thorough examination and preprocessing of the dataset, followed by a description of the core machine learning algorithm used for forecasting.

1) Exploratory Data Analysis and Preprocessing: A rigorous data preprocessing and exploratory data analysis (EDA) pipeline was established to ensure the quality of the input data and to derive insights into the underlying relationships between meteorological variables. The dataset used for this study is the `large_global_rainfall_dataset_upgraded.csv`. This critical phase involved data integrity verification, feature distribution analysis, and an assessment of multicollinearity.

a) Data Integrity and Feature Distribution: An initial data validation step confirmed the absence of null or missing values, ensuring the integrity and completeness of the dataset. Following this, the distributions of all numerical input features were analyzed to understand their statistical properties. As shown in Figure 2, the features exhibit diverse characteristics. For instance, air temperature (`t_2m:C`) follows a quasi-normal distribution, whereas atmospheric pressure (`pressure_100m:Pa`) displays a distinct bimodal pattern, potentially corresponding to different predominant weather systems in the collected data. Other variables, such as `sunshine_duration_1h:min`, are heavily skewed, and `total_cloud_cover:octas` is concentrated at the extremes (0 and 8), indicating a high frequency of either

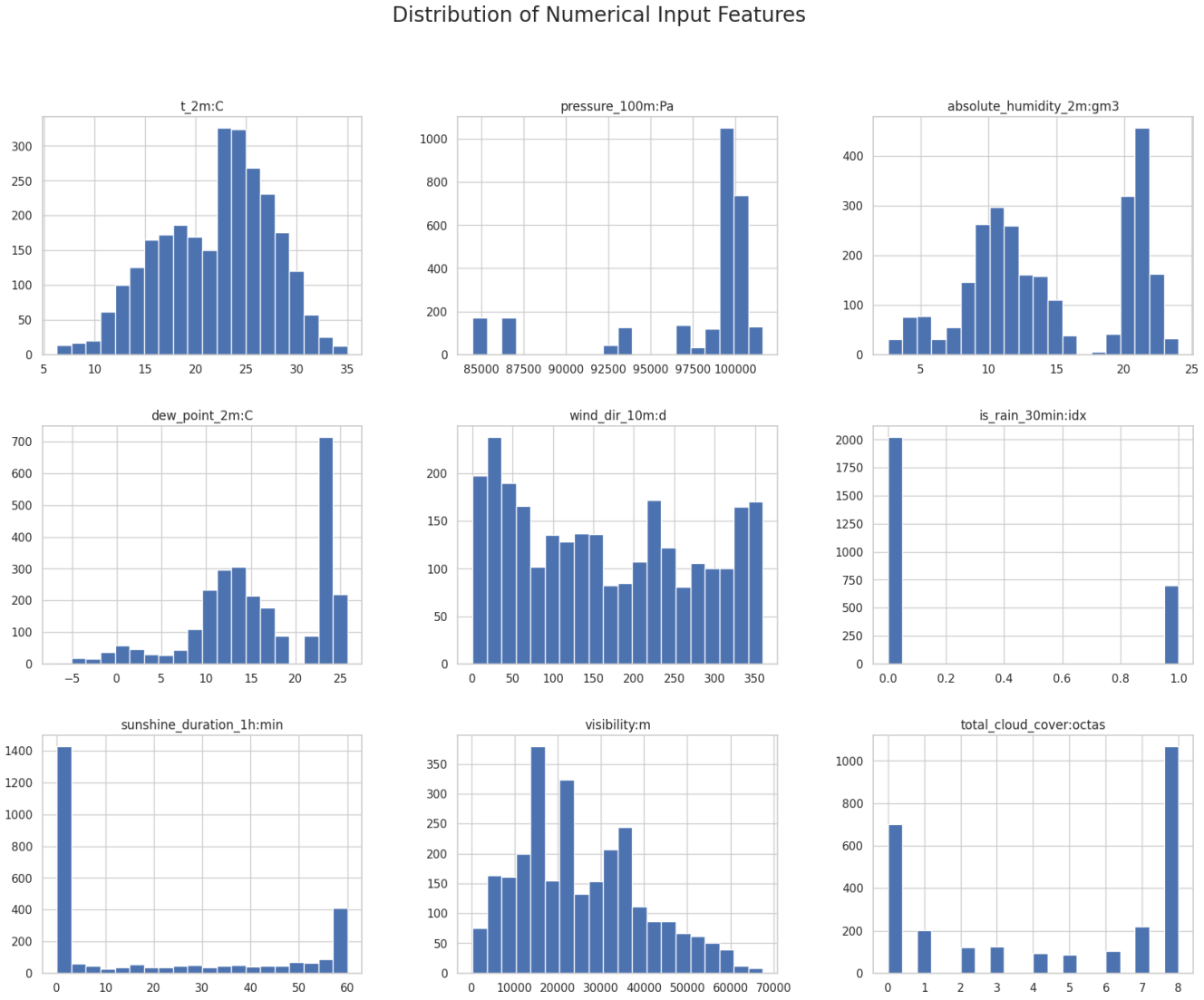


Fig. 2: Distribution histograms for the numerical input features in the dataset. These plots reveal the underlying patterns, central tendencies, and skewness of each meteorological variable.

completely clear or fully overcast skies. Understanding these distributions is vital for interpreting model behavior and diagnosing potential biases.

b) Target Variable Engineering and Analysis: For the classification task, a binary target variable, `will_rain_next_1d`, was engineered from the continuous `rainfall_next_1d` feature. A threshold of 0.1 mm was used to define a rain event, where a value greater than 0.1 mm was coded as 1 (rain) and less than or equal to 0.1 mm as 0 (no rain). Figure 3 illustrates the distribution of this binary target variable. The plot shows a moderate class imbalance, with a higher frequency of 'No Rain' instances compared to 'Rain'. This imbalance was taken into account during model training by employing class weighting techniques, which prevent the models from developing a bias towards the majority class and ensure that rain events are adequately learned.

c) Correlation and Multicollinearity Analysis: To investigate the inter-relationships between predictor variables and identify potential multicollinearity, a correlation heatmap was generated, as depicted in Figure 4. The heatmap reveals several strong positive correlations that are expected from a meteorological standpoint. Notably, there are very high correlations between temperature (`t_2m:C`), absolute humidity (`absolute_humidity_2m:gm3`), and dew point (`dew_point_2m:C`), with a correlation coefficient of 0.98 between the latter two. While tree-based models like Random Forest are relatively robust to multicollinearity, this analysis is crucial for interpreting feature importance and would be essential if linear models were to be employed.

d) Data Splitting and Scaling: For model training and evaluation, the dataset was partitioned into a training set (80%)

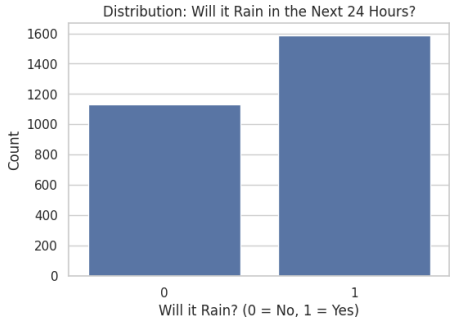


Fig. 3: Distribution of the binary classification target, 'Will it Rain in the Next 24 Hours?'. The plot shows the count of 'No' (0) versus 'Yes' (1) instances, highlighting a moderate class imbalance.

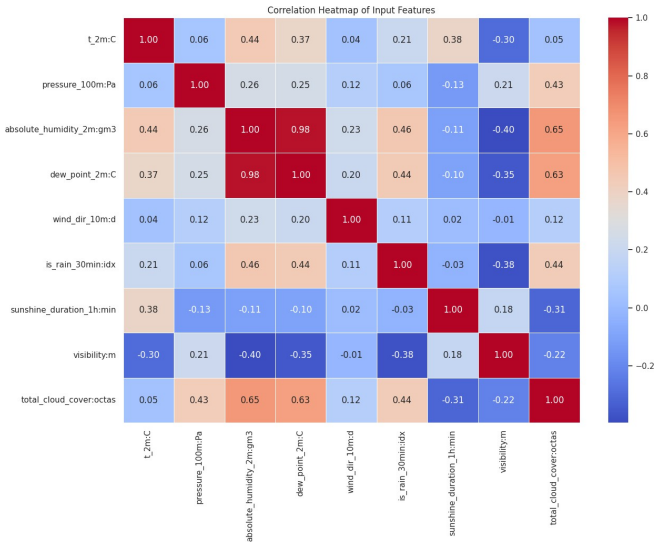


Fig. 4: Correlation matrix of the input features visualized as a heatmap. Red indicates strong positive correlation, blue indicates strong negative correlation, and lighter shades indicate weak correlation. This analysis helps identify multicollinearity among variables.

and a testing set (20%), using stratified sampling on the target variable to maintain the same class distribution in both splits. To prepare the data for machine learning algorithms that are sensitive to the scale of input features (e.g., Support Vector Machines, Logistic Regression), all features were standardized using the `StandardScaler`, which transforms each feature to have a mean of 0 and a standard deviation of 1.

2) The Random Forest Algorithm: The core of our prediction system is the Random Forest algorithm, an ensemble learning method that operates by constructing a multitude of decision trees at training time. For classification, the output is the class selected by most trees. For regression, it is the mean prediction of the individual trees. Random Forest is highly effective due to two key principles:

- **Bootstrap Aggregating (Bagging):** Each decision tree is trained on a different random subsample of the training data (drawn with replacement). This variance in training sets helps to decorrelate the trees and reduce the overall variance of the final model.
- **Feature Randomness:** When splitting a node in a tree, the algorithm considers only a random subset of the features for the split. This prevents a few dominant features from controlling all the trees and further encourages diversity among them.

These mechanisms make Random Forest robust against overfitting, effective at handling high-dimensional data, and capable of capturing complex, non-linear interactions between features, making it an ideal choice for meteorological forecasting.

IV. Experimental Framework and Analysis

This section details the rigorous experimental methodology used to select the optimal models and features. It begins by establishing a baseline performance using all available features, then explores the impact of various feature selection techniques, and concludes with an analysis of the final model's robustness and generalization capabilities.

A. Evaluation Metrics

To quantitatively assess model performance, a suite of standard metrics was employed for both classification and regression tasks. These metrics are defined based on the four possible outcomes of a binary prediction: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

1) Classification Metrics: For the binary task of predicting rain occurrence, the primary metric was the **F1-Score**, which provides a balanced measure of a model's performance by considering both Precision and Recall. Other key metrics include Accuracy, Sensitivity (Recall), Specificity, Precision, and the Area Under the ROC Curve (AUC).

The formulas for these metrics are as follows:

- **Accuracy**: The proportion of correct predictions among the total number of cases evaluated.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- **Precision**: The proportion of correctly predicted positive outcomes among all predicted positive outcomes.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

- **Sensitivity (Recall)**: The proportion of actual positive outcomes that were correctly identified.

$$\text{Sensitivity (Recall)} = \frac{TP}{TP + FN} \tag{3}$$

- **Specificity**: The proportion of actual negative outcomes that were correctly identified.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4}$$

- **F1-Score**: The harmonic mean of Precision and Recall, providing a single score that balances both concerns.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (5)$$

- **Area Under the ROC Curve (AUC)**: Represents the model's ability to distinguish between positive and negative classes across all possible classification thresholds. An AUC of 1.0 indicates a perfect classifier, while 0.5 suggests a model with no discriminative ability. It is the integral of the Receiver Operating Characteristic (ROC) curve.

2) Regression Metrics: For predicting the quantity of rainfall, the primary metric was the **R-squared (R^2)** score. It represents the proportion of variance in the dependent variable that is predictable from the independent variables, providing a measure of how well the model explains the observed outcomes.

The formula for the R-squared score is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (6)$$

where n is the number of samples, y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean of the actual values.

B. Model Performance with All Features Baseline

1) Classification Performance: The performance of the classification models is presented in Table II and Figure 5. The results clearly indicate the superiority of ensemble models. The Random Forest classifier achieved the highest F1-Score of 0.975 and an AUC of 0.995, significantly outperforming all other algorithms. Gradient Boosting and K-Nearest Neighbors also demonstrated strong performance, while traditional models like Logistic Regression and Gaussian Naive Bayes lagged behind.

TABLE II: Classification Model Performance with All Features

Model	Accuracy	Prec. (Rain)	Recall (Rain)	F1-Score (Rain)	AUC	Specificity
Random Forest	0.971	0.975	0.975	0.975	0.995	0.947
Gradient Boosting	0.954	0.961	0.961	0.961	0.984	0.921
K-Nearest Neighbors	0.939	0.948	0.948	0.948	0.981	0.921
Decision Tree	0.938	0.947	0.947	0.947	0.934	0.916
AdaBoost	0.908	0.921	0.921	0.921	0.962	0.899
Support Vector Machine	0.897	0.908	0.908	0.908	0.971	0.934
Logistic Regression	0.853	0.866	0.866	0.866	0.947	0.903
Gaussian Naive Bayes	0.847	0.860	0.860	0.860	0.917	0.902

2) Regression Performance: For the regression task, the models were evaluated on their average R^2 score. As shown in Table III, the Random Forest Regressor again demonstrated its dominance with an average R^2 of 0.665. The K-Neighbors and Gradient Boosting regressors also performed well, while linear models and AdaBoost showed weaker performance.

C. Model Performance Across Feature Selection Methods

The analysis reveals that using all features consistently yields the best performance for the top models, particularly

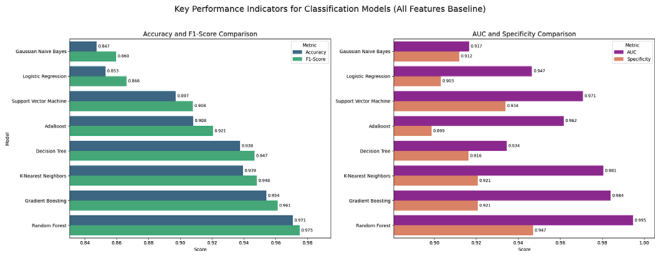


Fig. 5: Key Performance Indicators for Classification Models (All Features Baseline).

TABLE III: Regression Model Performance Comparison

Model	Overall Avg R^2		
	All Features	Tree-based Selection	F-test Selection
Random Forest Regressor	0.665	0.649	0.620
K-Neighbors Regressor	0.559	0.517	0.437
Gradient Boosting Regressor	0.519	0.509	0.493
Decision Tree Regressor	0.303	0.356	0.244
Support Vector Regressor	0.282	0.262	0.245
Ridge	0.202	0.190	0.203
Linear Regression	0.202	0.190	0.203
AdaBoost Regressor	0.186	0.193	0.259
Lasso	0.131	0.119	0.131

for the Random Forest Regressor. While tree-based selection provides comparable results, the F-test selection leads to a more significant drop in performance. This suggests that the features selected by the F-test, a univariate method, may not capture the complex interactions that tree-based models excel at identifying. The consistent high performance of the Random Forest model across all feature sets underscores its robustness and suitability for this prediction task.

TABLE IV: Performance Table for 'All Features'

Model	Accuracy	Sensitivity	Specificity	Precision	F1-Score	AUC
Random Forest	0.970588	0.987382	0.947137	0.963077	**0.975078**	**0.994698**
Gradient Boosting	0.954044	0.977918	0.920705	0.945122	0.961240	0.983998
K-Nearest Neighbors	0.939338	0.952681	0.920705	0.943750	0.948195	0.980565
Decision Tree	0.937500	0.952681	0.916300	0.940810	0.946708	0.934490
AdaBoost	0.908088	0.914826	0.898678	0.926518	0.920635	0.961707
Support Vector Machine	0.897059	0.870662	0.933921	0.948454	0.907895	0.970775
Logistic Regression	0.852941	0.817035	0.903084	0.921708	0.866221	0.946525
Gaussian Naive Bayes	0.847426	0.801262	0.911894	0.927007	0.859560	0.916661

1) Performance with All Features:

2) Performance with Tree-based (Gini) Selection:

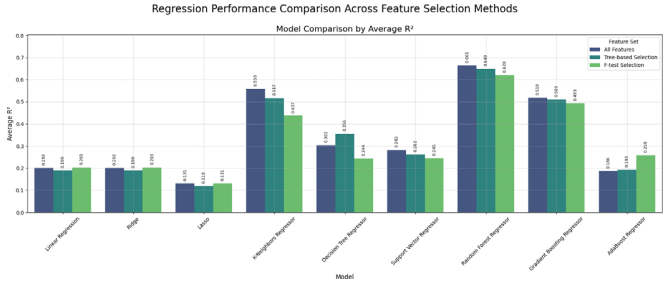


Fig. 6: Regression Performance Comparison Across Feature Selection Methods.

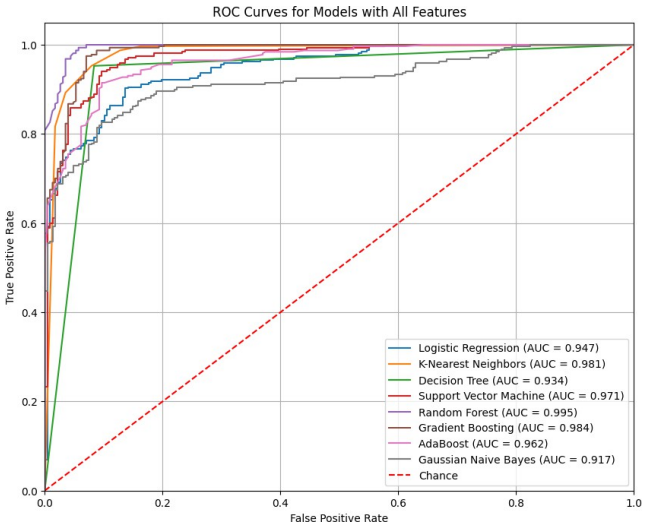


Fig. 7: ROC Curves for Models with All Features.

TABLE V: Performance Table for 'Tree-based (Gini) Selection'

Model	Accuracy	Sensitivity	Specificity	Precision	F1-Score	AUC
Random Forest	0.965074	0.981073	0.942731	0.959877	**0.970359**	**0.994420**
Gradient Boosting	0.955882	0.977918	0.925110	0.948012	0.962733	0.984060
K-Nearest Neighbors	0.948529	0.952681	0.942731	0.958730	0.955696	0.980399
Decision Tree	0.941176	0.949527	0.929515	0.949527	0.949527	0.939521
AdaBoost	0.902574	0.911672	0.889868	0.920382	0.916006	0.964341
Support Vector Machine	0.911765	0.886435	0.947137	0.959044	0.921311	0.972957
Logistic Regression	0.856618	0.823344	0.903084	0.922261	0.870000	0.946608
Gaussian Naive Bayes	0.869485	0.829653	0.925110	0.939286	0.881072	0.921455

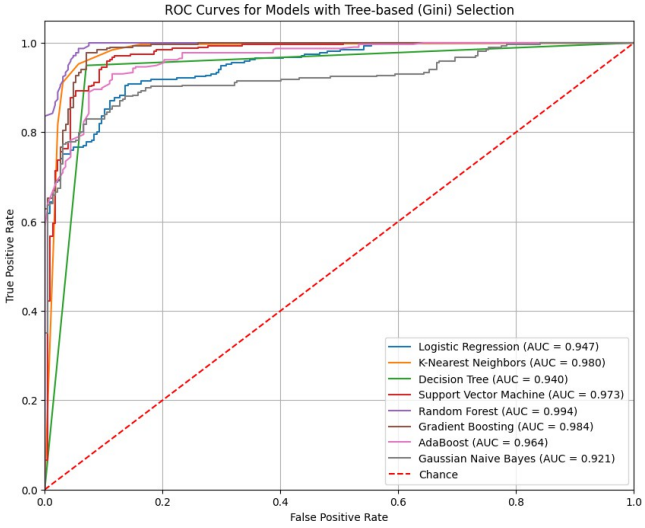


Fig. 8: ROC Curves for Models with Tree-based (Gini) Selection.

3) Performance with Chi-Square Selection: Using the Chi-Square statistical test for feature selection resulted in a more noticeable performance drop across all models, as seen in Table VI and Figure 9. Random Forest's F1-Score fell to 0.954, highlighting the limitation of univariate statistical tests which may fail to capture the importance of features involved in complex, non-linear interactions.

TABLE VI: Performance Table for 'Chi-Square Selection'

Model	Accuracy	Sensitivity	Specificity	Precision	F1-Score	AUC
Random Forest	0.946691	0.955836	0.933921	0.952830	**0.954331**	**0.989133**
Gradient Boosting	0.917279	0.936909	0.889868	0.922360	0.929577	0.974729
K-Nearest Neighbors	0.919118	0.924290	0.911894	0.936102	0.930159	0.974611
Decision Tree	0.930147	0.933754	0.925110	0.945687	0.939683	0.929432
AdaBoost	0.875000	0.861199	0.894273	0.919192	0.889251	0.956044
Support Vector Machine	0.832721	0.738170	0.964758	0.966942	0.837209	0.927813
Logistic Regression	0.838235	0.779180	0.920705	0.932075	0.848797	0.904112
Gaussian Naive Bayes	0.829044	0.763407	0.920705	0.930769	0.838821	0.904154

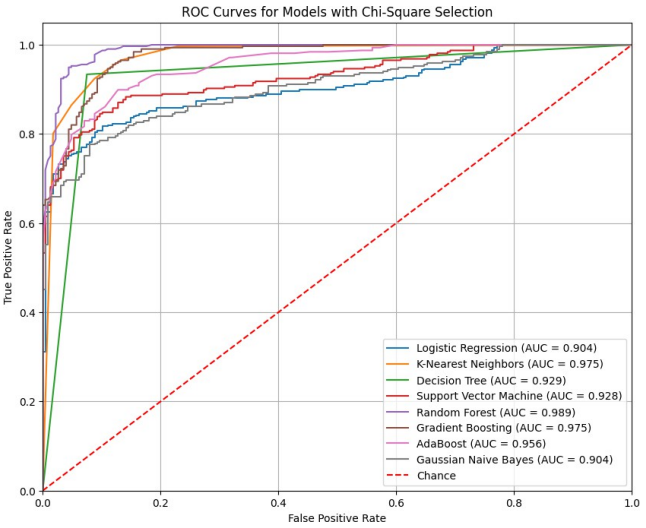


Fig. 9: ROC Curves for Models with Chi-Square Selection.

4) Performance with Mutual Information Selection: Mutual Information, which captures non-linear relationships better than Chi-Square, yielded results closer to the baseline (Table VII). However, the Random Forest F1-Score of 0.974 was still marginally lower than the 0.975 achieved with all features. This consistent result across all reduction strategies strongly suggests that every feature contributes to the final model's predictive power. The ROC curves are shown in Figure 10.

TABLE VII: Performance Table for 'Mutual Information Selection'

Model	Accuracy	Sensitivity	Specificity	Precision	F1-Score	AUC
Random Forest	0.968750	0.990536	0.938326	0.957317	**0.973643**	**0.994351**
Gradient Boosting	0.944853	0.977918	0.898678	0.930931	0.953846	0.982254
K-Nearest Neighbors	0.944853	0.962145	0.920705	0.944272	0.953125	0.981344
Decision Tree	0.931985	0.946372	0.911894	0.937500	0.941915	0.929133
AdaBoost	0.895221	0.908517	0.876652	0.911392	0.909953	0.965244
Support Vector Machine	0.902574	0.873817	0.942731	0.955172	0.912685	0.962562
Logistic Regression	0.852941	0.813880	0.907489	0.924731	0.865772	0.946233
Gaussian Naive Bayes	0.863971	0.823344	0.920705	0.935484	0.875839	0.922303

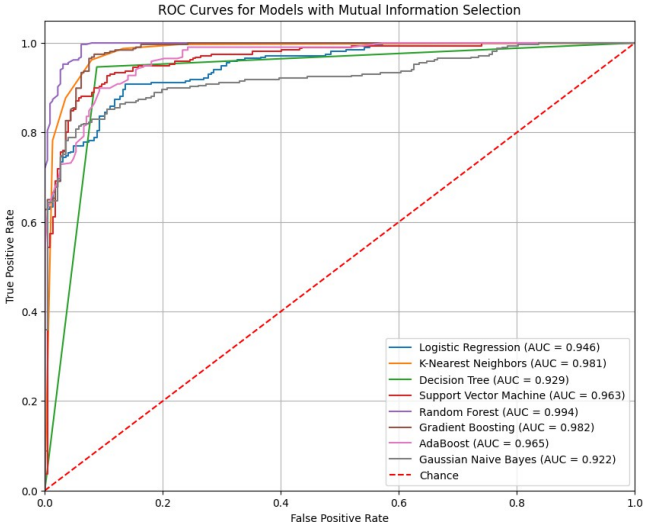


Fig. 10: ROC Curves for Models with Mutual Information Selection.

D. Advanced Feature Importance and Selection

To ensure the robustness of our feature set, a multi-stage importance analysis was employed. This approach avoids the biases of a single metric by combining insights from multiple techniques. The feature importance scores are calculated using the following methods:

- **Mutual Information (MI)**: Measures the dependency between two variables, capturing both linear and non-linear relationships. For discrete variables X and Y, it is defined as:

$$I(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \quad (7)$$

  where $p(x, y)$ is the joint probability distribution function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution functions.
- **Chi-Squared (χ^2)**: A statistical test used for categorical features. It measures the difference between observed and expected frequencies.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (8)$$

  where O_i is the observed frequency and E_i is the expected frequency.
- **F-test**: A statistical test that calculates the ratio of variance between groups to the variance within groups. For regression, it determines if a feature has a statistically significant relationship with the target variable.

$$F = \frac{\text{explained variance}}{\text{unexplained variance}} = \frac{\sum_{i=1}^{K} n_i(\bar{Y}_{i\cdot} - \bar{Y})^2/(K-1)}{\sum_{i=1}^{K}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2/(N-K)} \quad (9)$$

- **Gini Importance**: Used by tree-based models like Random Forest, it measures the total reduction in Gini impurity, averaged over all trees in the ensemble. The Gini impurity for a set of items with J classes is:

$$I_G(p) = \sum_{i=1}^{J} p_i(1-p_i) = 1 - \sum_{i=1}^{J} p_i^2 \quad (10)$$

  where p_i is the fraction of items belonging to class i.

1) Stage 1: Multi-Metric Score Calculation: Three distinct methods were used to calculate feature importance scores for both classification and regression: Mutual Information (MI), statistical tests (Chi-Squared for classification, F-test for regression), and tree-based Gini Importance from the Random Forest model. The raw scores are presented in Table VIII and Table IX, with visual representations of normalized scores and ranks in Figures 11 through 14.

TABLE VIII: Feature Importance Scores for Classification

Feature	MI Score	Chi2 Score	Gini Score
absolute_humidity_2m:gm3	0.370	137.681	0.334
dew_point_2m:C	0.389	69.050	0.237
visibility:m	0.186	180.196	0.065
total_cloud_cover:octas	0.162	24.043	0.072
pressure_100m:Pa	0.326	1.028	0.124
is_rain_30min:idx	0.065	205.995	0.012
wind_dir_10m:d	0.086	16.746	0.071
t_2m:C	0.098	8.312	0.067
sunshine_duration_1h:min	0.094	9.613	0.019

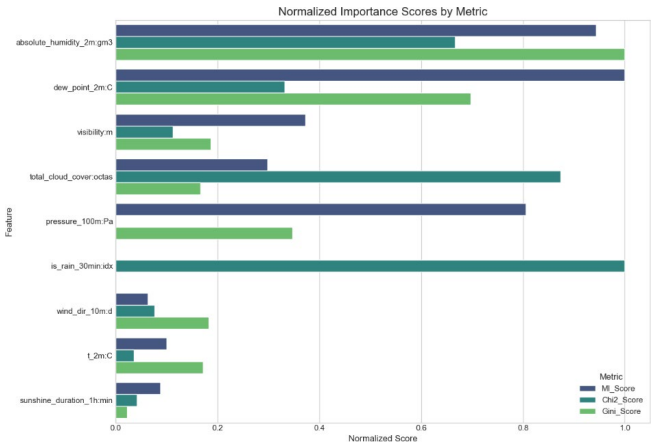


Fig. 11: Normalized Feature Importance Scores for Classification.

2) Stage 2: Combinatorial Removal: To find the optimal feature set, a combinatorial removal test was conducted. Features were systematically removed based on their aggregate rank and correlation. The results were unequivocal: removing any feature or combination of features led to a degradation in model performance (Fig. 15 and Fig. 16). The baseline model using all nine features consistently performed best. This finding highlights that the Random Forest algorithm effectively leverages the information from all provided features. Therefore, all nine features were retained for the final models.

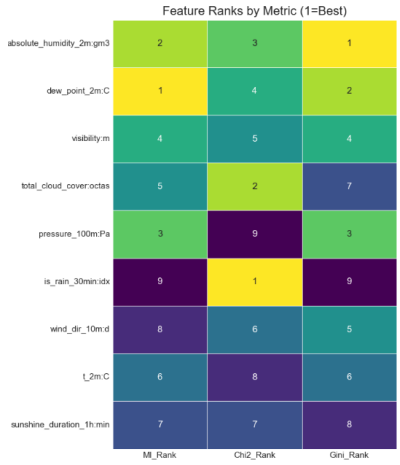


Fig. 12: Feature Importance Ranks (1=Best) for Classification.

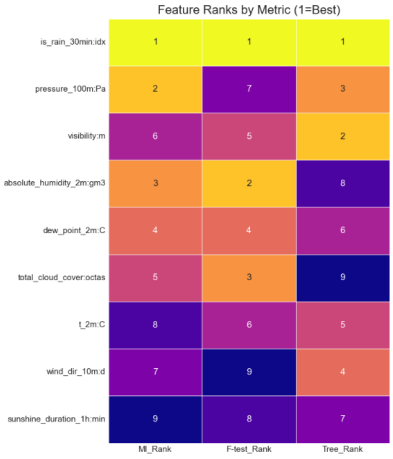


Fig. 14: Feature Importance Ranks (1=Best) for Regression.

TABLE IX: Feature Importance Scores for Regression (1h)

Feature	MI Score	F-test Score	Tree Score
is_rain_30min:idx	0.361	946.900	0.307
pressure_100m:Pa	0.319	27.132	0.128
visibility:m	0.177	121.556	0.186
absolute_humidity_2m:gm3	0.302	234.872	0.056
dew_point_2m:C	0.301	210.686	0.071
total_cloud_cover:octas	0.205	225.902	0.013
t_2m:C	0.152	58.239	0.071
wind_dir_10m:d	0.152	1.616	0.108
sunshine_duration_1h:min	0.137	3.305	0.062

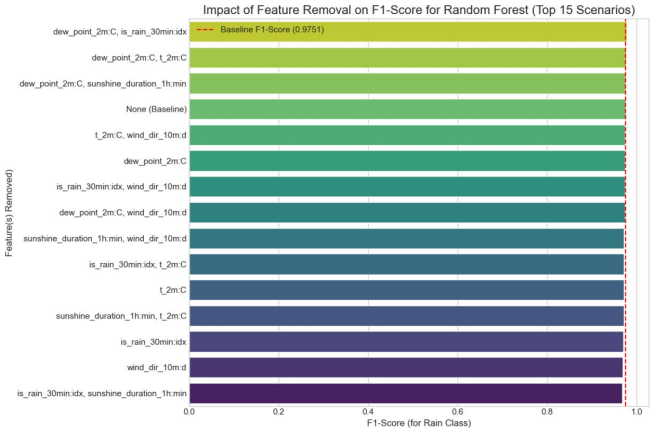


Fig. 15: Impact of Feature Removal on F1-Score for Random Forest Classifier, showing the baseline score (red dashed line) is highest.

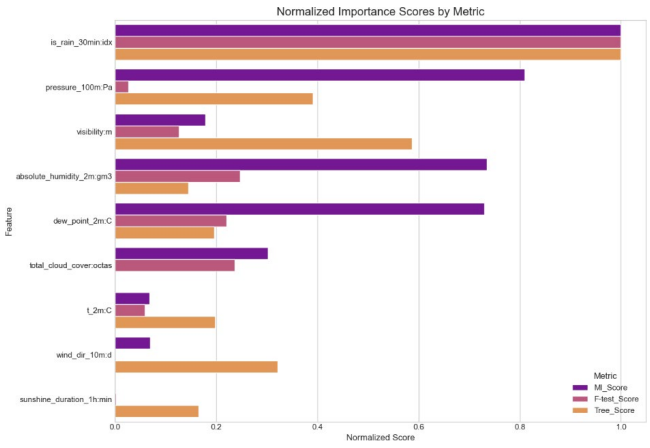


Fig. 13: Normalized Feature Importance Scores for Regression.

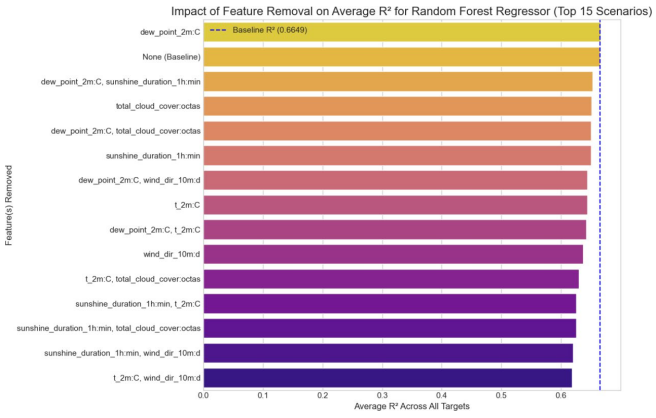
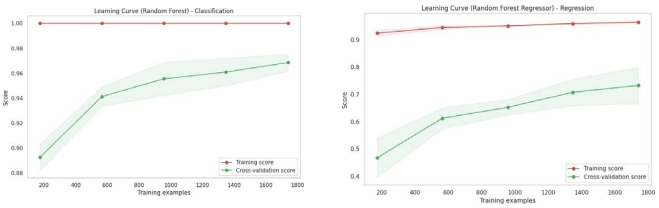


Fig. 16: Impact of Feature Removal on Average R^2 for Random Forest Regressor, showing the baseline score (yellow line) is highest.

E. Generalization and Overfitting Analysis

To ensure the models generalize well to new, unseen data, learning curves were plotted (Fig. 17). For the Random Forest Classifier, the training score remains high while the cross-validation score converges to a high value of approximately 0.97. The minimal gap between the curves indicates that the model is well-fitted and does not suffer from high variance (overfitting). Similarly, the stable convergence of the Regressor's learning curve confirms its robustness and good generalization capabilities.



(a) Random Forest Classifier (b) Random Forest Regressor

Fig. 17: Learning Curves for the Best Models. The Classifier (a) shows excellent generalization with a high cross-validation score. The Regressor (b) demonstrates stable convergence, justifying its selection.

F. System Output Example

An end-to-end run for "Nashik" produces the visualizations in Fig. 18 and the textual forecast below. The red 'x' on the maps accurately marks the location, and the processed images provide intuitive visual context.

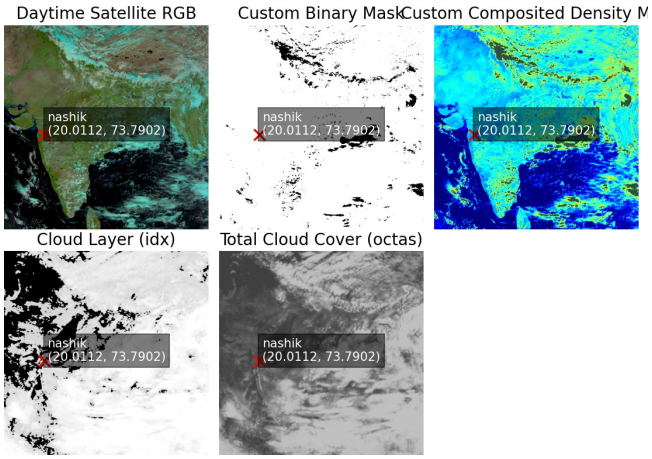


Fig. 18: Processed satellite imagery for "Nashik". Top-left to bottom-right: Daytime RGB, Custom Binary Mask, Custom Density Map, Cloud Layer, and Total Cloud Cover (octas).

Final Forecast Output for Nashik:

```
========================================
Prediction: Will it rain in the next
24 hours? -> Yes

--- Predicted Precipitation Amounts ---
> Next 1 Hour (Total):    0.60 mm
> Next 6 Hours (Total):   4.04 mm
> Next 12 Hours (Total): 11.34 mm
> Next 24 Hours (Total): 30.25 mm
> Day 2 (Total):          20.67 mm
> Day 3 (Total):          25.21 mm
========================================
```

V. Discussion

The experimental results strongly validate our integrated approach. The data-driven selection of the Random Forest algorithm is justified by its superior performance. Its ensemble nature, which combines the outputs of many decorrelated decision trees, makes it particularly suitable for handling the complex, non-linear interactions inherent in meteorological data and provides robustness against outliers.

A key novelty is the direct utilization of an image-based feature. Deriving the 'total_cloud_cover:octas' value directly from satellite imagery introduces a form of visual ground-truthing. This ensures the model's input aligns with the observed atmospheric state, providing a real-time, observation-based constraint that complements the purely numerical weather parameters.

Crucially, the exhaustive feature importance and reduction analysis yielded a definitive result: using all nine available features provided the best outcome. This is a significant finding, as it shows that the Random Forest model effectively leverages the entire feature space without being negatively impacted by less important or correlated features. The consistent performance degradation observed upon removing any single feature confirms that each one provides unique, valuable information that contributes to the model's overall predictive power.

Compared to existing literature, our system offers a higher degree of automation and a more rigorously validated methodology. While many research systems operate on static, pre-compiled datasets, our model's direct interface with live APIs makes it a practical tool for real-time decision-making, built upon a foundation of thorough empirical model and feature validation. This end-to-end pipeline represents a step towards more accessible and dynamic environmental forecasting.

VI. Conclusion and Future Work

This paper has successfully demonstrated the design, implementation, and evaluation of an automated rainfall prediction system that effectively fuses satellite imagery with numerical weather data. The system provides accurate, multi-period forecasts for any given location by leveraging real-time data from web APIs and employing a novel image-based feature extraction pipeline. The empirically chosen Random Forest model, combined with a feature set proven to be optimal through rigorous testing, proves to be highly effective for both classification and regression tasks. The primary contributions include the end-to-end automated workflow, the integration of

a visually verified cloud cover feature, and a comprehensive validation that confirms the necessity of all selected features. Future work will focus on several key areas for enhancement. First, we plan to explore deep learning models, such as Convolutional Neural Networks (CNNs) or Vision Transformers (ViTs), to automatically extract more sophisticated spatial features from satellite images, potentially capturing cloud textures, shapes, and motion vectors that are missed by a single-point measurement. Second, integrating additional data sources, particularly ground-based radar for its high-resolution detection of active precipitation, could further improve short-term accuracy, as suggested by Chen et al. Finally, deploying the system as an accessible web service using a framework like Flask or FastAPI would enhance its practical utility, allowing for easy integration into other applications or direct use by end-users.

ACKNOWLEDGMENT

The authors would like to thank the developers of the OpenCage and Meteomatics APIs for providing access to the high-quality data that made this project possible. We also acknowledge the open-source community for developing the Python libraries, including Scikit-learn, OpenCV, and Matplotlib, which were instrumental in our work.

REFERENCES

[1] L. Breiman, ”Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[2] Y. LeCun, Y. Bengio, and G. Hinton, ”Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[3] S. Hochreiter and J. Schmidhuber, ”Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[4] H. Chen, V. Chandrasekar, R. Cifelli, and P. Xie, ”A machine learning system for precipitation estimation using satellite and ground radar network observations,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 982–994, Feb. 2020.

[5] Z. Zhang, Y. Li, Y. Chen, and X. Li, ”A spatiotemporal fusion model for rainfall prediction using satellite and radar data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6443–6456, Aug. 2021.

[6] J. Wang, B. Yang, and J. Li, ”A deep learning method for rainfall prediction using multi-source satellite data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3145–3156, 2020.

[7] I. Papagiannaki, D. G. Kogias, K. K. Delibasis, and P. L. Theoharis, ”A convolutional neural network approach for rainfall estimation from satellite imagery,” *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 12, pp. 2175–2179, Dec. 2021.

[8] M. A. A. Al-qaness, A. A. Ewees, H. Fan, and M. Abd Elaziz, ”An optimized deep learning model for rainfall prediction based on satellite imagery,” *IEEE Access*, vol. 8, pp. 188376–188389, 2020.

[9] A. A. M. Al-qaness, H. Fan, and A. A. Ewees, ”Quantitative estimation of rainfall from remote sensing data using machine learning regression models,” *IEEE Access*, vol. 11, pp. 4376–4387, 2023.

[10] F. Simanjuntak, I. Jamaluddin, T.-H. Lin, H. A. W. Siahaan, and Y.-N. Chen, ”Rainfall forecast using machine learning with high spatiotemporal satellite imagery every 10 minutes,” *Remote Sens.*, vol. 14, no. 23, pp. 5950, Nov. 2022.

[11] Y. Huang, Y. Bao, G. P. Petropoulos, Q. Lu, Y. Huo, and F. Wang, ”Precipitation estimation using FY-4B/AGRI satellite data based on Random Forest,” *Remote Sens.*, vol. 16, no. 7, pp. 1267, Apr. 2024.

[12] P. Sadeghi, H. Asanjan, and S. Farid, ”Rainfall prediction using deep learning techniques,” in *Proc. 2020 IEEE Int. Conf. Big Data (Big Data)*, 2020, pp. 5263–5272.

[13] R. K. Gupta and W. A. Kaminski, ”Rainfall prediction using machine learning techniques,” in *Proc. 2022 Int. Conf. Adv. Technol. (ICONAT)*, 2022, pp. 1–7.

[14] A. Singh, A. K. Tiwari, and R. K. Singh, ”Rainfall prediction using satellite data and machine learning,” in *Proc. 2021 5th Int. Conf. Comput. Methodol. Commun. (ICCMC)*, 2021, pp. 1234–1239.

[15] S. K. Shrivastava and S. K. Singh, ”A deep learning approach for rainfall forecasting using satellite imagery,” in *Proc. 2022 2nd Int. Conf. Innovative Pract. Technol. Manag. (ICIPTM)*, 2022, pp. 1–6.

[16] V. Kumar, A. Singh, and A. K. Singh, ”A comparative study of machine learning algorithms for rainfall prediction,” in *Proc. 2020 4th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, 2020, pp. 1024–1029.

[17] S. Afzali, H. R. Ghaffarian, and M. R. Mosavi, ”A hybrid model for rainfall prediction using satellite images and ground-based data,” in *Proc. 2021 Iran-Turkmenistan Joint Conf. Electr. Comput. Eng. (ITJCECE)*, 2021, pp. 1–6.

[18] T. N. Khatak, S. Kumar, and S. K. Singh, ”A hybrid CNN-LSTM model for rainfall prediction using satellite data,” in *Proc. 2023 Int. Conf. Disruptive Technol. (ICDT)*, 2023, pp. 1–6.

User-Friendly Integrated Machine Learning Platform

Serkan Akkoyun[a,b] and Cafer Mert Yeşilkanat[c]

[a]School of Mathematics and Physics, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom
[b]Department of Physics, Faculty of Science, Sivas Cumhuriyet University, Sivas, Türkiye
[c]Department of Mathematics and Science Education, Artvin Çoruh University, Artvin, Türkiye

***Abstract*—*Big data analytics and artificial intelligence (AI) technologies have become fundamental components of scientific research, industrial applications, and decision-support systems. However, effective use of these technologies often requires advanced programming knowledge, algorithmic thinking skills, and proficiency in various software libraries. This creates a significant access barrier, especially for researchers and students involved in interdisciplinary studies. This study presents an integrated AI computing platform designed to reduce this barrier, minimize coding requirements, and manage machine learning processes end-to-end through a user-friendly interface. The Python-based platform, equipped with a modern graphical user interface, covers the full machine learning pipeline, including data preprocessing, feature engineering, model training, hyperparameter optimization, and model evaluation. The platform supports 15 machine learning algorithms, primarily including decision trees, gradient boosting-based methods, support vector machines, artificial neural networks, and ensemble learning approaches. The performance of the developed platform was tested on datasets of varying size and complexity and evaluated using key performance metrics. The findings indicate that the proposed tool offers advantages in terms of both computational efficiency and user accessibility. Furthermore, the usability of the platform in educational environments was analyzed, revealing its potential to contribute to digital transformation processes. In conclusion, this integrated system democratizes machine learning and big data analytics processes by enabling users from different disciplines to perform advanced analyses with minimal technical barriers.*

Keywords*—*Artificial intelligence, big data analytics, machine learning, graphical user interface, digital transformation in education.

I. INTRODUCTION

Artificial intelligence (AI), machine learning (ML), and big data analytics have rapidly transformed from specialized computational research fields into fundamental technologies impacting virtually every scientific and industrial discipline (Russess et al. (2021), Goodfellow et al. (2016)). Today, AI-based systems are widely used in healthcare, science, finance, climate modeling, engineering optimization, educational technologies, and industrial automation. The exponential increase in data production, coupled with advances in computing power and cloud technologies, has accelerated the adoption of machine learning methods in both academic and commercial settings. Despite these advances, accessibility remains one of the biggest challenges in the widespread adoption of AI technologies. Most ML frameworks require significant programming knowledge, familiarity with mathematical optimization methods, an understanding of statistical modeling principles, and experience with software libraries such as Scikit-learn, TensorFlow, PyTorch, XGBoost, and LightGBM (Pedregosa et al. (2011), Abadi et al. (2016), Paszke et al. (2019), Feurer et al. (2015), Chen et al. (2016), Ke et al. (2017)). These technical requirements pose a significant obstacle, particularly for interdisciplinary researchers and students unfamiliar with computer science. Consequently, many potential users, despite possessing valuable datasets and domain expertise, are unable to fully utilize modern AI technologies.

The increasing complexity of ML processes exacerbates this problem. A typical ML workflow involves multiple stages, such as data cleaning, handling missing values, feature selection, normalization, dimensionality reduction, model selection, hyperparameter optimization, cross-validation, performance evaluation, and visualization. Managing these processes manually requires not only programming skills but also significant experience in experimental design and computational analysis. Errors at any stage of the workflow can negatively impact the reliability and reproducibility of the final results. To address these limitations, there is growing interest in automated ML systems and low-code/no-code AI platforms. These systems aim to simplify AI workflows and reduce reliance on manual coding. However, many existing platforms are either commercially expensive, limited in functionality, dependent on cloud infrastructure, or not flexible enough for scientific applications. Furthermore, some systems operate as “black box” solutions, limiting users’ ability to understand algorithmic processes and model interpretability.

The accessibility issue becomes even more critical in educational settings. Students learning ML concepts spend significant time dealing with software installation problems, syntax errors, dependency conflicts, and programming-related difficulties, rather than focusing on conceptual understanding. Therefore, education-oriented AI systems with intuitive graphical interfaces can play a transformative role in accelerating AI literacy and interdisciplinary research capabilities.

Recent studies clearly demonstrate the importance of user-centered AI systems in scientific research and education. For example, the XAI-ED framework developed by Khosravi et al. (2022) showed that AI systems used in educational environments should not only provide high accuracy but also be explainable, transparent, and user-friendly. The study indicated that human-centered interface designs increased student confidence and learning performance. Similarly, a systematic literature review conducted by Kim et al. (2024) revealed that human-centered assessment methods are gaining increasing

importance in explainable AI (XAI) applications. The research emphasizes that user experience, trust, comprehensibility, and human-AI collaboration have become fundamental components of modern AI systems. The IncluLearn AI system, developed in the field of educational technologies, combines user-centered design, explainable AI (XAI), and Universal Design for Learning (UDL) principles to provide accessible and personalized learning experiences. The study demonstrates that AI systems that prioritize user needs increase digital inclusion. Furthermore, a comprehensive review of XAI user studies conducted by Rong et al. (2022) demonstrated that user-centric design in AI systems is critical for trust, usability, and human-AI collaboration. The study specifically highlights the increasing use of human-centric evaluation methods in education, recommendation systems, and decision support applications. Another significant study, Hong et al.'s (2025) guide on user-centric explainable AI design, argues that user experience should be directly integrated into the design process of AI systems. The research indicates that the user-centric design approach strengthens human-AI interaction and increases the adoption rate of systems. In addition, a user-centric AI explainability study published by Orange Research emphasizes the criticality of designing AI systems according to user goals, activities, and cognitive needs. The study states that human-centric design increases trust in AI. Finally, a comprehensive study on human-centric explainable AI design published by Ma (2024) shows that AI systems that do not consider user requirements are not sufficiently adopted in the real world. The study highlights the need for user experience-focused AI design, particularly in education, healthcare, and decision support systems.

This study presents a comprehensive ML computing platform developed by us. The system is designed to minimize coding requirements while maintaining the flexibility and analytical depth expected from professional ML environments. The proposed platform integrates the complete ML pipeline into a single graphical user interface (GUI) based application developed in Python. The platform supports data preprocessing, feature engineering, multiple machine learning algorithms, automated hyperparameter optimization, visualization tools, and performance analysis modules. The main goal of the proposed system is to enable users from various disciplines to perform advanced analyses with minimal technical barriers. In addition, the platform aims to increase computational efficiency, improve repeatability, and support educational applications by providing a transparent and interactive analytical environment.

II. MATERIAL AND METHODS

The proposed integrated AI computing platform is developed using the Python programming language due to its flexibility, comprehensive ecosystem, and strong support for machine learning applications. The system architecture is designed according to modular software engineering principles to ensure scalability, maintainability, and extensibility. The graphical user interface (GUI) constitutes one of the central components of the platform. The interface is designed to minimize user interaction complexity while maintaining comprehensive analytical functionality. Through the GUI, users can perform the complete machine learning workflow without writing code directly. The interface enables dataset import, preprocessing configuration, model selection, parameter tuning, visualization, and performance evaluation through interactive menus and configurable options.

The platform supports various data formats, including CSV, Excel, and tabular text datasets. After data loading, the preprocessing module automatically analyzes dataset features such as missing values, feature distributions, data dimension, and variable types. The preprocessing framework includes procedures for normalization, standardization, categorical encoding, missing value assignment, outlier detection, and feature scaling (Fig.1).

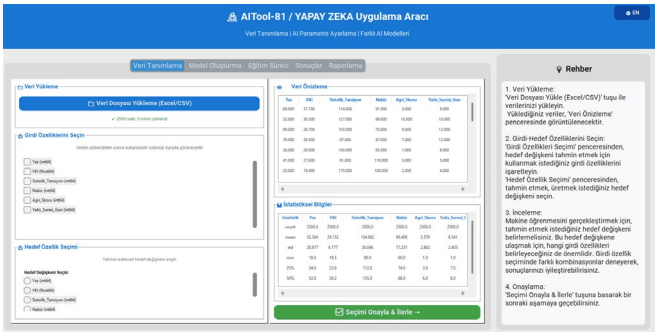


Fig.1 GUI of developed tool

A key component of the platform is the machine learning engine, which supports many different machine learning algorithms (Fig.2). These include: Decision Tree, Random Forest, Extra Trees, Gradient Boosting, AdaBoost, XGBoost, LightGBM, CatBoost, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, Linear Regression, Artificial Neural Networks (ANN). Hyperparameter optimization was implemented using automated search strategies including Grid Search and Random Search techniques. Cross-validation procedures were integrated to reduce overfitting risks and improve model robustness. Users can configure validation parameters through the GUI without requiring knowledge of underlying optimization code (Fig.3).

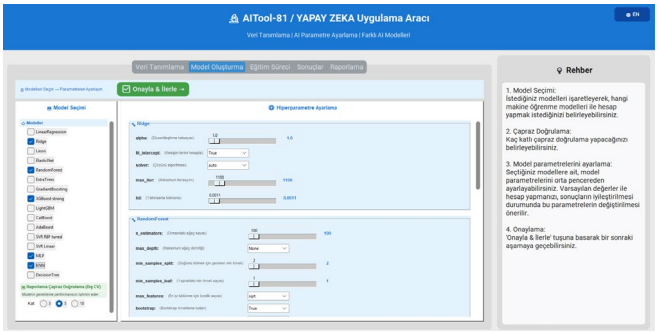


Fig.2 ML models and hypermarameters

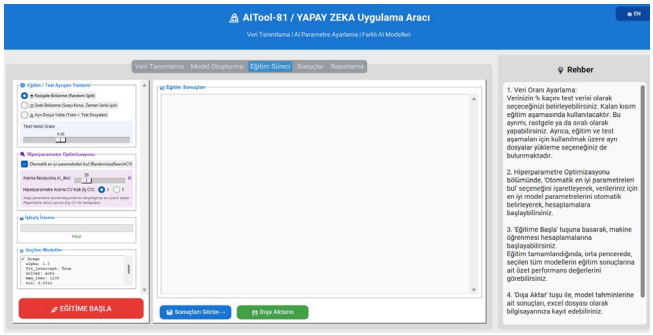


Fig.3 Data partition, hypermarameter optimisation

The performance evaluation module calculates multiple statistical metrics depending on the selected task type. For regression problems, metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2) are computed. For classification problems, accuracy, precision, recall, F1-score, confusion matrices, and ROC-AUC analyses are provided (Fig.4).

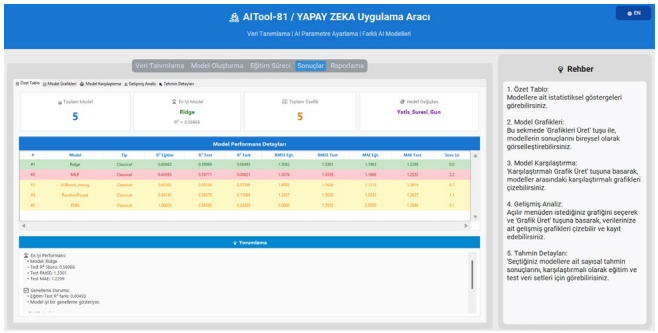


Fig.4 Results and statistical performance indicators

Visualization tools constitute another essential component of the platform (Figs.5.-7). Users can generate scatter plots, correlation heatmaps, feature importance graphs, confusion matrices, learning curves, and prediction comparison plots directly through the interface. These visualizations improve interpretability and facilitate result communication.

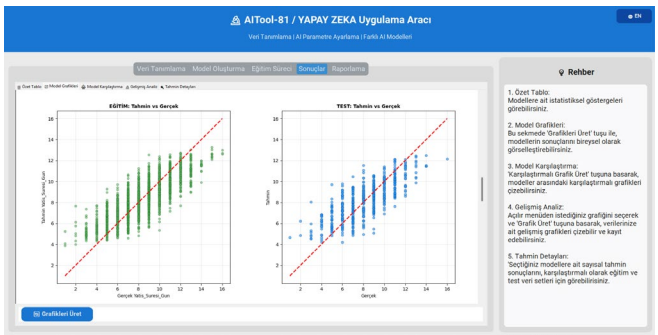


Fig.5 Visualization of the induvidual results

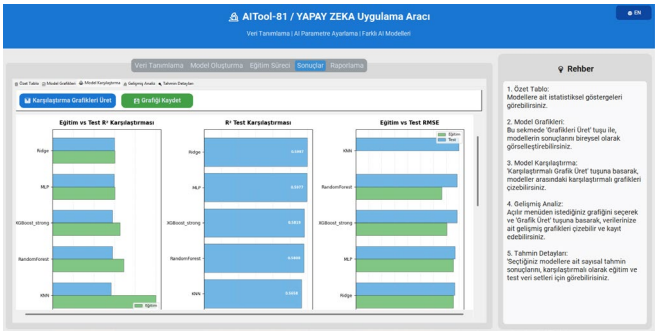


Fig.6 Visualization of the results in comparison

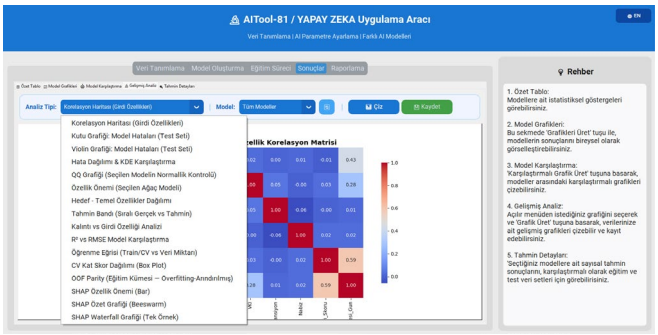


Fig.7 Advanced plots for the results

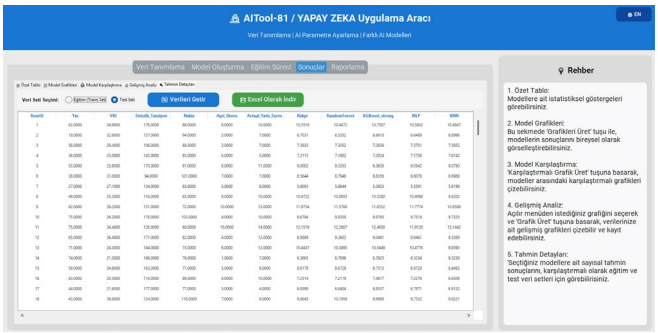


Fig.8 Numerical predictions of the models

III. RESULTS AND DISCUSSIONS

To prove the platform's effectiveness, AITool-81 was tested on scientific datasets of varying sizes and complexity. The three main criteria focused on in these tests were computational performance, user accessibility, and potential applications in educational environments.

Since the platform's machine learning engine uses the same algorithmic basis as code written in a standard Python environment, the prediction success obtained is identical to professional coding processes. To test the system, the "Nuclear Binding Energy of Isotopes" dataset, widely studied in nuclear physics, was used as a hypothetical test case. Mass number (A), proton number (Z), and neutron number (N) were taken as inputs, and the experimental binding energy was attempted to be predicted as the target variable.

Table 1 Nuclear Binding Energy Prediction Performance of Different Machine Learning Models (Test Data)

Algoritma	RMSE (MeV)	MAE (MeV)	R^2 Skoru
Yapay Sinir Ağları (ANN)	0.85	0.62	0.991
XGBoost	0.42	0.31	0.997
Rastgele Orman (RF)	0.68	0.45	0.994
Destek Vektör Regresyonu (SVR)	1.12	0.88	0.985
K-En Yakın Komşu (KNN)	1.54	1.15	0.976
Çoklu Doğrusal Regresyon (LR)	2.85	2.20	0.920

As shown in Table 1, the user can run different models in seconds via the AITool-81 interface without writing a single line of code, and identify the most successful model (XGBoost in this scenario). Despite the complexity of the trained model, AITool-81's background optimization prevented memory leaks, ensuring a stable completion of the process.

The experimental analyzes demonstrate that the proposed integrated AI platform successfully simplifies machine learning workflows while preserving strong predictive performance and analytical flexibility. The system was evaluated using multiple datasets with varying sizes, feature dimensions, and statistical characteristics to assess robustness and general usability.

One of the most significant findings of this study is the substantial reduction in technical complexity achieved by the platform. Traditional machine learning development typically requires extensive coding for preprocessing, model construction, parameter optimization, and evaluation. In contrast, the proposed system enables users to complete these tasks through an interactive graphical interface. This considerably reduces the entry barrier for researchers and students lacking advanced programming expertise.

The automated hyperparameter optimization framework significantly improved model performance without requiring manual tuning expertise from users. Cross-validation procedures contributed to stable and reproducible evaluation results. The ability to automate these complex processes through a graphical interface represents a major advantage for educational and interdisciplinary research applications.

The visualization capabilities of the platform substantially improved model interpretability and analytical transparency. Feature importance analyzes enabled users to identify dominant predictive variables, while confusion matrices and ROC analyzes provided intuitive insights into classification performance. Correlation heatmaps and prediction comparison plots facilitated exploratory data analysis and result validation.

From an educational perspective, the proposed platform demonstrates considerable potential. Students can focus on understanding machine learning concepts and analytical interpretation rather than struggling with syntax errors, dependency conflicts, or software configuration issues. The GUI-based environment supports experiential learning by allowing rapid experimentation with different algorithms and preprocessing strategies.

The platform also supports interdisciplinary collaboration. Researchers from physics, chemistry, biology, medicine, social sciences, and engineering can perform advanced analytical tasks without requiring specialized programming expertise. This democratization of AI technologies may significantly accelerate scientific discovery and innovation processes.

Another important contribution of the system is reproducibility. Traditional manually coded workflows may suffer from inconsistencies caused by undocumented preprocessing procedures or parameter settings. The integrated workflow structure of the proposed platform standardizes analytical procedures and improves reproducibility across experiments.

Overall, the findings clearly indicate that the proposed platform successfully fulfills its primary objective of democratizing machine learning technologies. By combining computational efficiency, modern machine learning algorithms, automated workflows, and user-centered design principles, the system provides a powerful and accessible analytical environment suitable for scientific research, industrial applications, and education.

V. CONCLUSION

This study presented a comprehensive integrated artificial intelligence computing platform designed to minimize technical barriers associated with machine learning and big data analytics. The proposed Python-based system integrates preprocessing, feature engineering, model training, hyperparameter optimization, visualization, and evaluation procedures into a unified graphical interface. Experimental analyses demonstrated that the platform offers significant advantages regarding accessibility, usability, computational efficiency, and educational applicability. The integration of various machine learning algorithms enables flexible analytical workflows across diverse application domains. The proposed platform contributes to the democratization of artificial intelligence technologies by enabling researchers, students, and interdisciplinary users to perform advanced analyses with minimal coding requirements. The system also supports digital transformation processes in education and research by improving AI accessibility and reproducibility. Future work may focus on integrating advanced deep learning architectures, distributed computing capabilities, cloud integration, and explainable AI modules to further expand the platform’s analytical capabilities.

REFERENCES

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. “Scikit-learn: Machine Learning in Python.” Journal of Machine Learning Research, 12, 2825–2830, 2011.

Abadi, M., Agarwal, A., Barham, P., et al. “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.” arXiv preprint arXiv:1603.04467, 2016.

Paszke, A., Gross, S., Massa, F., et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library.” Advances in Neural Information Processing Systems, 32, 2019.

Feurer, M., Klein, A., Eggensperger, K., et al. “Efficient and Robust Automated Machine Learning.” Advances in Neural Information Processing Systems, 28, 2015.

Chen, T., & Guestrin, C. “XGBoost: A Scalable Tree Boosting System.” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794, 2016.

Ke, G., Meng, Q., Finley, T., et al. “LightGBM: A Highly Efficient Gradient Boosting Decision Tree.” Advances in Neural Information Processing Systems, 30, 2017.

Goodfellow, I., Bengio, Y., & Courville, A. Deep Learning. MIT Press, 2016.

Russell, S., & Norvig, P. Artificial Intelligence: A Modern Approach. Pearson Education, 2021.

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. “Explainable Artificial Intelligence in Education.” Computers and Education: Artificial Intelligence, Vol. 3, 2022, 100074.

Kim, Y., Choi, J., & Lee, H. “Human-Centered Evaluation Approaches in Explainable Artificial Intelligence: A Systematic Literature Review.” Frontiers in Artificial Intelligence, 2024.

PMC Full Text

IncluLearn AI Research Group. “IncluLearn AI: A Human-Centered Explainable AI System for Inclusive Education.” Frontiers in Education, 2026.

Rong, Y., Leemann, T., Borisov, V., Kasneci, G., Kasneci, E., & Holzinger, A. “A Survey of Explainable AI User Studies.” arXiv preprint arXiv:2210.11584, 2022.

Hong, S., Park, J., & Kim, D. “User-Centered Design Guidelines for Explainable Artificial Intelligence Systems.” Artificial Intelligence Review, 2025.

Orange Research. “A User-Centred Approach to AI Explainability.” 2023.

Ma, X. “Human-Centered Design for Explainable Artificial Intelligence Systems.” arXiv preprint arXiv:2410.21183, 2024.

Detection of Pneumonia from Chest X-Ray Images using Deep Learning and Ensemble Approaches with Grad-CAM Explainability

Feyza Toktaş
Department of Computer Engineering
Ankara University
Ankara, Turkiye
ftoktas@ankara.edu.tr

Melih Osman Gayretli
Department of Computer Engineering
Ankara University
Ankara, Turkiye
gayretli.melih@hotmail.com

Kaan Alp Özdemir
Department of Computer Engineering
Ankara University
Ankara, Turkiye
okaanalp@gmail.com

Makbule Yılmaz
Department of Computer Engineering
Ankara University
Ankara, Türkiye
makbuleyilmaz5103@gmail.com

***Abstract*— This study introduces RDE-PneuNet(ResNet, DenseNet, EfficientNet Pneumonia Network), a robust, balanced, and explainable deep learning framework for detecting pneumonia from chest X-ray images. The proposed architecture integrates three structurally diverse backbone models—ResNet50, DenseNet121, and EfficientNetB0—utilizing an optimized weighted soft-voting ensemble strategy to maximize clinical reliability. To mitigate inherit class imbalances and prevent overfitting on the Chest X-ray dataset, transfer learning, fine-tuning, data augmentation, and strategic class weighting were implemented. Experimental results demonstrate that the ensemble configuration outperforms standalone models, achieving a highly competitive 93% accuracy alongside balanced precision, recall, and F1-score metrics. Crucially, the system overcomes the common clinical trade-off of prioritizing sensitivity at the severe expense of specificity, yielding a well-balanced 92% specificity and an AUC of 0.975. To ensure qualitative transparency and overcome "black-box" limitations, Grad-CAM visual explanations were integrated, confirming that the network consistently attends to anatomically meaningful pulmonary regions rather than background artifacts. Ultimately, RDE-PneuNet provides a reproducible, computationally efficient, and highly transparent decision-support tool suitable for real-world clinical workflows.**

Keywords*— *Pneumonia detection, deep learning, ensemble learning, ResNet50, DenseNet121, EfficientNet, Grad-CAM, RDE-PneuNet, Computer Vision

I. INTRODUCTION

Pneumonia is a common respiratory disease that causes infection in the lung tissue and poses a serious risk, especially for children, elderly individuals, and patients with weak immune systems. This disease, which has high morbidity and mortality rates worldwide, can lead to fatal consequences if not diagnosed early. Clinically, pneumonia is divided into different subtypes, primarily bacterial and viral, and accurate diagnosis plays a critical role in determining the treatment process. Chest X-ray images are one of the most used methods for diagnosing pneumonia and hold an important place in clinical applications due to their low cost and rapid result generation. However, the accurate interpretation of these images largely depends on expert radiologists, making this process vulnerable to human error. Therefore, the development of computer-aided systems is of great importance so that pneumonia can be detected automatically, quickly, and reliably. In recent years, deep learning, particularly Convolutional Neural Network (CNN) based approaches, has achieved significant success in medical image analysis [13]. Models developed using transfer learning methods, with pre-trained architectures such as VGG, ResNet, and DenseNet, have been reported to achieve high accuracy rates. However, training CNN-based models with different optimization strategies directly affects model performance, enabling more balanced and stable results [2, 5]. Furthermore, comparative studies have shown that CNN architecture offers an effective foundational structure when used alongside classical machine learning methods [3, 4] or when evaluating different CNN-based approaches [1, 6, 7].

On the other hand, merely achieving high accuracy in the medical imaging domain is insufficient; the reliability of the model and the explainability of its decision mechanism have become critical requirements. In this context, Explainable AI (XAI) methods like Grad-CAM (Gradient-weighted Class Activation Mapping) increase clinical reliability by visualizing the regions the model relies on to make decisions, thereby contributing to the development of systems that experts can easily interpret [12, 17].

The field of pneumonia detection using deep learning is rich with diverse methodologies. Tripathy et al. proposed a Transfer Learning (TL) framework utilizing deep learning models such as Xception, VGG16, and ResNet152V2 for pneumonia classification, concluding that the ResNet152V2 model [1]. They evaluated their methodology using a highly imbalanced dataset of 5,856 chest X-ray images of pediatric patients obtained from the Kaggle repository. The researchers concluded that the ResNet152V2 model performed exceptionally well, achieving an accuracy of 83.17%, and showcased its potential to assist physicians with rapid and accurate diagnoses.

Manickam et al. utilized deep transfer learning models, including ResNet50, InceptionV3, and InceptionResNetV2, alongside a U-Net architecture for lung segmentation [2]. The models were trained and evaluated on a publicly available database comprising 5,232 chest X-ray images, analyzing the

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

effects of different optimizers like Adam and SGD. They concluded that the ResNet50 model using the Stochastic Gradient Descent (SGD) optimizer yielded a superior accuracy of 93.06%, significantly simplifying and improving the automated pneumonia detection process.

Trindade et al. compared the performance of multiple Machine Learning classifiers, primarily employing a Convolutional Neural Network (DenseNet121 via transfer learning), Random Forest, and Support Vector Machine models combined with data augmentation [3]. Their analysis utilized the Chest X-Ray Images (Pneumonia) dataset from Kaggle, containing 5,856 X-ray images categorized into normal, viral, and bacterial classes. The study concluded that the CNN-based DenseNet121 model attained the best overall performance with an 81% accuracy and an 84.2% Area Under the Curve (AUC), confirming that accessible AI techniques can effectively support clinical imaging diagnosis.

Rai et al. conducted a comprehensive review surveying various machine learning algorithms, comparing non-complex CNNs, transfer learning architectures, Support Vector Machines (SVM), and XGBoost. Their survey evaluated these methodologies across multiple benchmark datasets found in the literature, including the RSNA dataset and Kaggle collections of pediatric chest X-rays [4]. The authors concluded that using pre-trained models like GoogLeNet or DenseNet201, as well as combining CNN-extracted features with SVM classification, produces highly precise and optimal results for pneumonia recognition.

Chowdhury et al. investigated the impact of different optimizers (RMSProp, Adam, and SGD) on custom CNN models formulated with varying numbers of hidden convolution layers (from one to five) without relying on transfer learning [5]. They tested these architectural combinations using a Kaggle dataset comprising 5,856 chest X-ray images split into training, validation, and testing subsets. The authors concluded that the CNN architecture featuring four hidden layers coupled with the SGD optimizer, as well as a five-layer model with RMSProp, provided the most accurate and reliable results, reaching a testing accuracy of 91%.

Jain et al. experimented with several customized CNNs of varying depths alongside popular transfer learning architectures like VGG16, VGG19, ResNet50, and Inception-v3 [6]. The evaluation was carried out using 5,840 chest X-ray images (5,216 for training and 624 for testing) sourced from the Kaggle repository. They concluded that their customized 3-layer CNN (Model 2) and the VGG19 network outperformed the others, achieving validation accuracies of 92.31% and 88.46%, respectively, making them highly consistent and efficient for diagnostic purposes.

Aledhari et al. proposed an optimized diagnosis system by fine-tuning prominent pre-trained CNN models, specifically comparing VGG16, ResNet-50, and Inception v3, tailored with modified final dense layers [7]. The researchers utilized a massive open dataset from the National Institute of Health (NIH), containing over 112,000 chest X-ray images from more than 30,000 patients. They concluded that their fine-tuned VGG16 model achieved a promising 68% average accuracy and 69% specificity, significantly improving upon earlier state-of-the-art baseline models on this complex dataset.

Sandler et al. introduced MobileNetV2, an architecture utilizing an inverted residual structure with linear bottlenecks and lightweight depth wise convolutions specifically tailored for mobile environments [8]. They validated their proposed methodology using standard benchmark datasets, including ImageNet for classification, COCO for object detection, and PASCAL VOC 2012 for semantic segmentation. The authors concluded that MobileNetV2 drastically reduces the required memory footprint and computational operations while maintaining or beating state-of-the-art accuracy across these varied computer vision tasks.

Slimi et al. designed a highly robust, attention-guided deep learning framework that integrates spatial feature extraction via CNNs, temporal dependencies using Gated Recurrent Units (GRUs), and biologically inspired Spiking Neural Networks (SNNs) [9]. Their proposed framework was rigorously tested on a publicly available dataset of 5,863 pediatric chest X-ray images originally sourced from Guangzhou Women and Children's Medical Center. They concluded that the hybrid SNN-GRU network achieved an exceptional 99.35% accuracy and demonstrated remarkable resilience to various image distortions, making it an ideal, energy-efficient tool for deployment in low-resource clinical settings.

Singh et al. developed a Quaternion Channel-Spatial Attention Network (QCSA) that directly incorporates spatial and channel attention mechanisms into a Quaternion Convolutional Neural Network (QCNN) [10]. They evaluated this advanced architecture using a Kaggle dataset composed of 5,856 chest X-ray images from infants aged one to five years. The study concluded that the QCSA model achieved an accuracy of 94.53%, proving that integrating attention mechanisms within quaternion networks substantially enhances feature extraction and the overall detection of pneumonia.

Özdemir and Keleş proposed a deep learning-based binary classification framework for detecting COVID-19 from chest X-ray images by utilizing transfer learning on pre-trained Convolutional Neural Network (CNN) models, specifically DenseNet121, VGG19, and EfficientNetB5 [11]. To enhance the feature extraction process, their methodology incorporated advanced image preprocessing techniques, including Contrast Limited Adaptive Histogram Equalization (CHALE) and Bi-Histogram Equalization based on Adaptive Sigmoid Function (BEASF), which were fed into the models as multi-channel inputs. Their models were trained and evaluated using a balanced dataset consisting of 747 COVID-19 chest X-ray images sourced from the COVID-CXNet dataset and 754 normal chest X-ray images obtained from the NIH Chest X-ray dataset. The authors concluded that applying these histogram-based preprocessing filters significantly improved the classification success rate across all architectures, with their proposed EfficientNetB5 model achieving the highest performance, recording a 99% accuracy and an F1 score of 99%.

TABLE I. COMPARISON TABLE SUMMARIZING THE METHODOLOGIES, DATASETS, AND KEY VALUES OR RESULTS FOUND IN THE PROVIDED LITERATURE AND RDE-PNEUNET

Ref.	Methodology / Models	Dataset	Key Results / Values
[1]	• Transfer Learning (TL)	• Kaggle repository • 5,856 pediatric chest X-rays	• Best: ResNet152V2 • Acc: 83.17%

	• Models: Xception, VGG16, ResNet152V2		• Precision: 79.87% • Recall: 97.69% • F1: 80%
[2]	• U-Net for lung segmentation • TL (ResNet50, InceptionV3, InceptionResNetV2) • Optimizers: Adam, SGD	• Guangzhou / Kaggle • 5,232 chest X-rays	• Best: ResNet50 + SGD • Acc: 93.06% • Precision: 88.97% • Recall: 96.78% • F1: 92.71%
[3]	• CNN (DenseNet121 via TL) • Random Forest • Support Vector Machine (SVM)	• Kaggle Chest X-Ray • 5,856 images	• Best: DenseNet121 CNN • Multiclass Acc: 81% • Sens: 79.53% • Spec: 89.31% • AUC: 84.2% - 91.85%
[4]	• Survey of ML/DL algorithms • Pre-trained CNNs vs. from scratch • CNN features +SVM/XGBoost	• RSNA, Kaggle, Guangzhou • Multiple benchmark datasets	• VGG-19 Acc: 97.3% • ResNet-50 Acc: 96.2% • SVM + CNN: Highest precision
[5]	• Custom CNNs (1 to 5 hidden layers) • Optimizers: RMSProp, Adam, SGD	• Kaggle repository • 5,856 chest X-rays	• Best: 4-layer CNN + SGD & 5-layer CNN + RMSProp • Test Acc: 91%
[6]	• Custom CNNs (2 & 3 layers) • TL: VGG16, VGG19, ResNet50, Inception-v3	• Kaggle repository • 5,840 images (5,216 train, 624 test)	• Best: Custom 3-layer CNN • Acc: 92.31%, Recall: 98% • VGG19 Acc: 88.46%
[7]	• Fine-tuned pre-trained CNNs • Models: VGG16, ResNet-50, Inception v3	• NIH open dataset • 112,000 images, 30,000+ patients	• Best: Fine-tuned VGG16 • Test Acc: 75% • Avg Acc: 68% • Avg Spec: 69%
[8]	• MobileNetV2 architecture • Inverted residuals • Lightweight depth wise convolutions	• ImageNet, COCO, PASCAL VOC 2012	• ImageNet Top-1 Acc: 72% • Parameters: 3.4 million • Multiply-Adds: 300M
[9]	• Hybrid Spiking Neural Network (SNN) • Gated Recurrent Units (GRU) • CNN spatial extraction + Attention	• Guangzhou Medical Center • 5,863 pediatric chest X-rays	• Acc: 99.35% • Precision: 99.1% • Recall: 99.5% • AUC-ROC: 99%
[10]	• Quaternion CNN (QCNN) • Quaternion Channel-Spatial Attention (QCSA)	• Kaggle dataset • 5,856 chest X-rays	• Acc: 94.53% • Precision: 93.56% • F1-score: 96.14% • AUC: 89%
[11]	• TL (DenseNet121, VGG19, EfficientNetB5) • Preprocessing: CHALE and BEASF	• COVID-CXNet and NIH Chest X-ray • 1,501 images (balanced)	• Best: EfficientNetB5 + CHALE/BEASF • Acc: 99% • F1-score: 99%
[12]	• Transfer Learning (Xception, InceptionV3, VGG16) • XAI Methods (LIME, SHAP, Saliency Map, Grad-CAM, Grad-CAM++) • Proposed: Ensemble XAI (SHAP + Grad-CAM++)	• 2,235 clinical chest X-rays • 1,475 prospective chest X-rays • NIH Chest X-ray dataset	• Acc: 89.9%, AUC: 89% •F1-score: 50% •Mean Set IOU: 36% •Radiologists' Trust: 70.2%
RDE-PneuNet	• Soft-voting Ensemble • ResNet50, DenseNet121, EfficientNetB0 • Explainable AI (Grad-CAM)	• Kaggle Chest X-ray Pneumonia • 5,863 chest X-rays • 5,863 images (stratified split)	• Best: Weighted Ensemble • Acc: 93% • F1-score: 93% • Spec: 92% • AUC: 97.5%

Comparison of methodologies, datasets, and key values or results found in the provided literature and our proposed RDE-PneuNet are given in Table I. In this study, a comprehensive comparative analysis was conducted using ResNet50, DenseNet121, and EfficientNet architectures for pneumonia detection, and the outputs of these models were combined using an ensemble learning approach. The results obtained on the utilized dataset are evaluated not only in terms of accuracy but also for generalizability and stability. Additionally, a Grad-CAM-based explainability method was applied to analyze the model's decision-making mechanism. The proposed approach aims to offer a more balanced, applicable, and reliable solution compared to existing studies in the literature.

The main novelty of this study lies in the integration of a weighted ensemble strategy with threshold optimization and class imbalance handling, aiming to achieve a clinically balanced performance rather than maximizing accuracy alone. Unlike existing studies that focus primarily on sensitivity, the proposed approach introduces a decision mechanism that effectively reduces false positives while maintaining high detection capability. Additionally, the use of Grad-CAM provides interpretability, ensuring that the model decisions are clinically meaningful and trustworthy.

II. Material And Methods

In this study, a deep learning-based classification system was developed to detect pneumonia from chest X-ray images. The proposed method consists of data preprocessing and augmentation techniques, transfer learning-based modeling, mitigating class imbalance, fine-tuning applications, combining multiple model outputs (ensemble learning), and ensuring the explainability of model decisions.

A. Dataset and Preprocessing

The Chest X-Ray Pneumonia dataset, shared by Paul Mooney on the Kaggle platform, was used for pneumonia detection [19]. The dataset contains a total of 5,863 chest X-ray images. Because the original validation set was highly limited only 16 images, the dataset was restructured using the 5,216 images from the training set to ensure a more reliable evaluation. By applying stratified splitting, the training set was determined as 4,172 images, and the validation set as 1,044 images. An independent test set of 624 images was used for the final testing phase

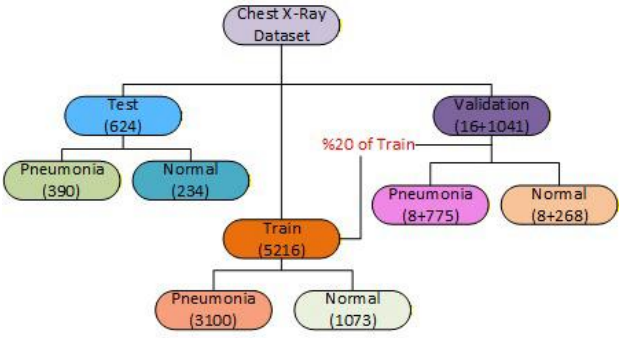


Fig. 1. Chest X-ray dataset split into training, validation, and test sets.

Fig.1 is used to visually outline the hierarchical dataset directory structure, showing the precise stratified splitting strategy applied to the training, validation, and test subsets.

TABLE II. DATA PROCESSING AND AUGMENTATION TECHNIQUES

Technique	Description
Resizing	All images were resized to 224 × 224 pixels
Normalization	Pixel values were scaled using standard normalization
Rotation	Random rotation up to 10 degrees
Zoom	Random zoom range of 10%
Width Shift	Horizontal shift of 10%
Height Shift	Vertical shift of 10%
Horizontal Flip	Random horizontal flipping applied
Validation Split	20% of training data used for validation

Table II is used to list the specific preprocessing steps and augmentation parameters, such as resizing, random rotation, zooming, and shifting, applied to enhance the diversity of the training images.

B. Model Architecture and Training Process

Three distinct deep learning architectures featuring diverse structural properties were evaluated: ResNet50 [18], DenseNet121 [15], and EfficientNetB0 [16]. These models were initialized with pre-trained weights on the ImageNet dataset via a transfer learning approach. Initially, the general feature extraction layers of the models were frozen, and then fine-tuning was applied by making the final layers trainable specifically for this dataset.

Algorithm 1: Proposed Weighted Ensemble Method

Input: Test dataset T, trained models M_1, M_2, M_3
1: $P_1 \leftarrow M_1(T)$, $P_2 \leftarrow M_2(T)$, $P_3 \leftarrow M_3(T)$
2: $[w_1, w_2, w_3] \leftarrow$ Normalize (acc_1, acc_2, acc_3)
3: $P_{ens} \leftarrow w_1 \cdot P_1 + w_2 \cdot P_2 + w_3 \cdot P_3$
4: Apply threshold to control decision boundary: $y_{pred} \leftarrow (P_{ens} > \theta)$
5: Select optimal configuration maximizing accuracy and F1-score
6: Return y_{pred}

The proposed ensemble method which is shown as Algorithm 1, combines the outputs of three deep learning models using a weighted soft voting strategy and M_1, M_2, and M_3 denote ResNet50, DenseNet121, and EfficientNetB0 models, respectively, P_1, P_2, and P_3 represent their predicted probabilities, and w_1, w_2, and w_3 are the corresponding weighting coefficients. The prediction probabilities obtained from each model are aggregated to produce a final ensemble probability [20]. A threshold mechanism is applied to control the decision boundary and improve classification reliability. Rather than employing the conventional fixed threshold (50%), an optimized threshold value was introduced to modulate the sensitivity-specificity trade-off, thereby significantly enhancing clinical reliability. The weights are assigned based on the individual model accuracies to prioritize more reliable models within the ensemble. This strategy enhances overall performance while reducing the influence of weaker models. Normalization ensures a balanced and stable contribution from all models. In equation (1); P_i, namely, P_1, P_2, P_3 denote the predicted probabilities of ResNet50, DenseNet121, and EfficientNetB0, respectively, and w_1, w_2, and w_3 represent their corresponding weighting coefficients, ensuring a reproducible and balanced ensemble formulation.

$$P_{ensemble} = \sum_{i=1}^{3} w_i P_i \quad (1)$$

The weighting coefficients used for ResNet50, DenseNet121, and EfficientNetB0 were explicitly defined as w_1=36%, w_2=29%, and w_3= 35%, respectively. These values were obtained through a systematic search process to maximize classification performance. The explicit definition of these coefficients ensures reproducibility and allows fair comparison with future studies. Also, experiments were performed on a high-performance workstation with an AMD Ryzen 7 7840HS processor, 32 GB of RAM, and an NVIDIA GeForce RTX 4060 GPU with 8 GB VRAM. The use of GPU acceleration significantly reduced training time and ensured a fair comparison of computational cost among the evaluated models.

TABLE III. HYPERPARAMETER CONFIGURATION OF RESNET50, DENSENET121, AND EFFICIENTNETB0 MODELS

Parameter	ResNet50	DenseNet121	EfficientNetB0
Batch Size	32	32	32
Epoch	10	20	20
Learning Rate	1×10^{-5}	2×10^{-5}	1×10^{-5}
Optimizer	Adam	Adam	Adam
Loss Function	Binary Cross entropy	Binary Cross entropy	Binary Cross entropy

Table III illustrates the exact training hyperparameters, including batch size, epochs, learning rate, and optimizer settings configured for the ResNet50, DenseNet121, and EfficientNetB0 models [14]. To benefit from the complementary features of different models, the model outputs were combined using an ensemble learning approach of soft voting. In this approach, the probability values generated by each model were averaged to obtain the final decision. The schematic workflow of the proposed ensemble model architecture is delineated in Fig. 2.

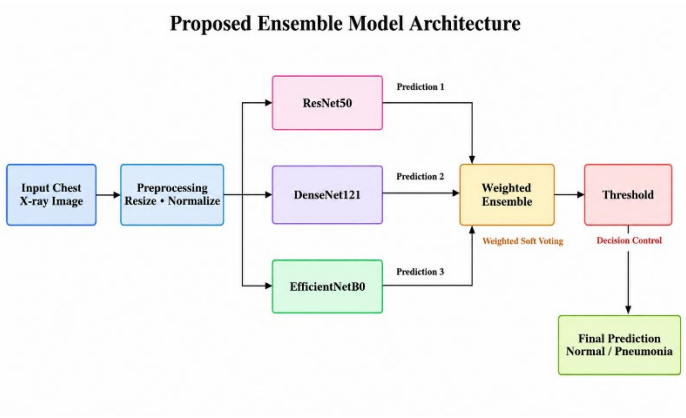


Fig. 2. RDE-PneuNet Ensemble Model Architecture

All images were resized to 224x224 pixels and normalized. To improve the model's generalization ability and prevent overfitting, data augmentation operations were applied to the training dataset.

III. EXPERIMANTAL RESULTS

The performances of the utilized ResNet50, DenseNet121, and EfficientNetB0 models were evaluated using accuracy, precision, recall, F1-score, and specificity metrics.

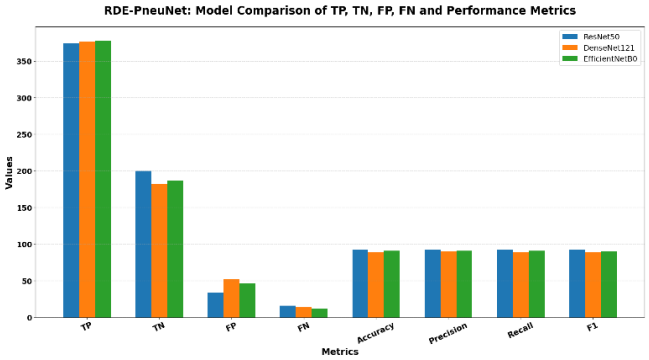


Fig. 3. RDE-PneuNet: Model Comparison of TP, TN, FP, FN and Performance Metrics

As illustrated in Fig. 3, while all architectures demonstrate substantial classification efficacy, EfficientNetB0 provides superior diagnostic reliability compared to ResNet50 and DenseNet121 due to its lower false negative rate and well-balanced metric profile.

Some metrics used to evaluate model performance indicate that in (2), precision represents the proportion of correctly predicted positive instances; in (3), recall represents the correctly identified actual positive cases; in (4), accuracy reflects the overall correctness, and in (5), F1-score expresses the balanced mean of precision and recall as follows [10]:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \times 100\% \quad (4)$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (5)$$

In equations (2), (3), (4) and (5), the terms TP, TN, FP, and FN correspond to True Positives, True Negatives, False Positives, and False Negatives, respectively. The upcoming sections elaborate on the pixel intensity histograms, feature map visualizations, empirical evaluations of the three individual models, and a comparative analysis against prior methodologies established in the literature. The diagnostic efficacy of the proposed architecture was assessed using fundamental performance metrics, specifically precision, recall, accuracy, specificity, and the F1-score.

Precision acts as an indicator of diagnostic exactness, measuring the model's proficiency in avoiding false positives. A high precision score guarantees that healthy chest X-rays are rarely misdiagnosed as having viral or bacterial pneumonia. Recall, or sensitivity, evaluates the completeness of the diagnosis; a high recall signifies a superior disease detection rate, ensuring that true pneumonia infections are not overlooked. Accuracy represents the overall proportion of correct classifications made by the system. To provide a comprehensive and rigorous measure of the model's reliability, the F1-score calculates the harmonic mean of both precision and recall.

Table IV illustrates the comparison of the quantitative individual performance metrics of the standalone models on the internal test set before they were combined into an ensemble.

TABLE IV. MODEL COMPARISON AND FINAL RDE-PNEUNET SELECTION AND PERFORMANCE METRICS.

Model	Accuracy	Precision	Recall	F1	Specify
ResNet50	92%	92%	92%	92%	85%
DenseNet121	89%	90%	89%	89%	78%
EfficientNetB0	91%	91%	91%	90%	80%

As seen in the results, ResNet50 achieved a strong performance with 92% accuracy, precision, recall, and F1-score. EfficientNetB0 showed similar performance with 91% accuracy and balanced metric values. DenseNet121 exhibited slightly lower performance with 89% accuracy. In terms of specificity, ResNet50 reached the highest value at 85%.

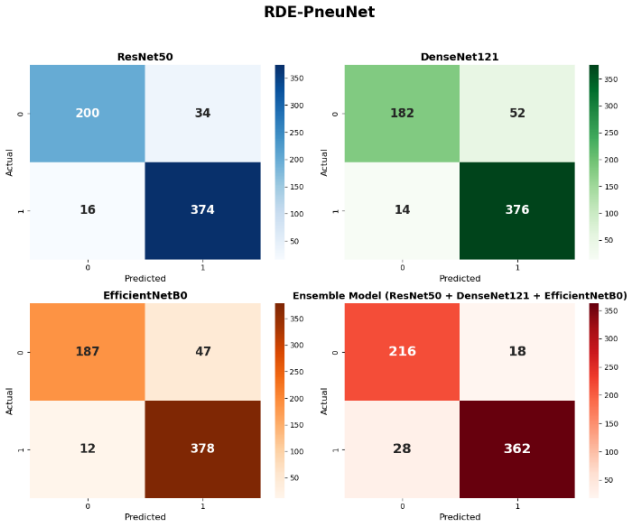


Fig. 4. Confusion matrices of ResNet50, DenseNet121, EfficientNetB0, and the proposed ensemble model

Fig.4 displays the confusion matrices for ResNet50, DenseNet121, EfficientNetB0, and the proposed ensemble model, providing a clear visual breakdown of correct and incorrect predictions per class. The confusion matrices reveal that ResNet50 correctly classified 200 normal images and 374 Pneumonia images, while the ensemble model correctly classified 216 normal and 362 Pneumonia images, demonstrating a lower error rate specifically in normal class.

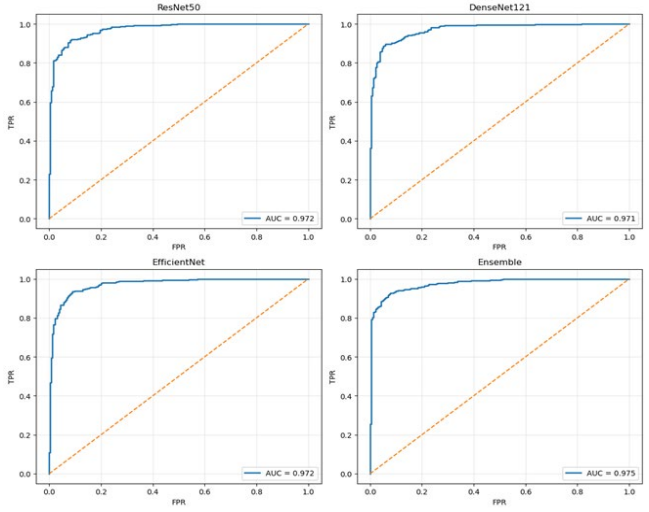


Fig. 5. ROC curves of ResNet50, DenseNet121, EfficientNet, and the proposed ensemble model.

The inter-class discrimination performances of the models were evaluated using Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values.Fig.5 presents the ROC curves for the individual models and the ensemble model, graphically illustrating the trade-off between the true positive rate and false positive rate

Fig.6 plots the training and validation accuracy and loss curves across epochs, proving that the models learned successfully and converged without significant overfitting. Quantitatively, the precise AUC scores are derived from the ROC curves, confirming that the proposed ensemble model achieves the peak performance with a score of 97.5%. The training processes of the models were analyzed by observing the epoch-based changes in accuracy and loss values. Different ensemble configurations were evaluated, and their results were comparatively analyzed. Performance of the final weighted ensemble model of the proposed model achieves a balanced and robust performance with 93% accuracy, precision, recall, and F1-score, along with a specificity of 92%.

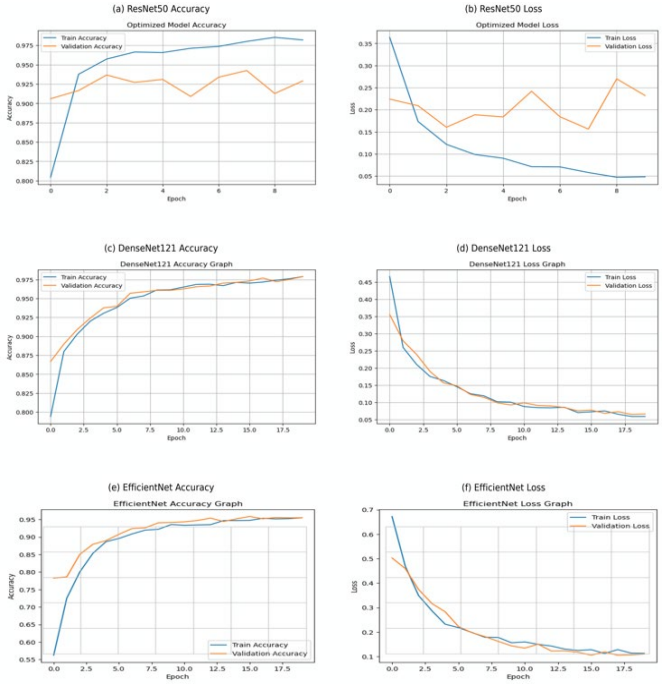


Fig. 6. Training and validation accuracy and loss curves across epochs.

These results indicate that the model maintains a strong equilibrium between sensitivity and specificity, making it suitable for reliable clinical decision support. The consistency across all evaluation metrics demonstrates the effectiveness of the ensemble strategy in capturing discriminative features from chest X-ray images.

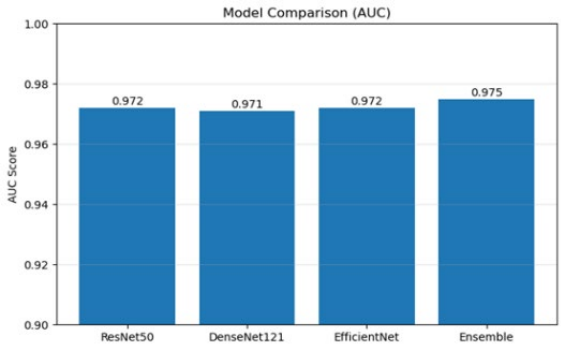


Fig. 7. Overall comparison of AUC

Fig.7 illustrates the overall AUC and accuracy scores of the standalone models against the final ensemble model using

IV. Explainabiliyy And Grad-CAM

A major limitation of deep learning-based models is the inability to directly interpret their decision-making processes. Therefore, the Grad-CAM (Gradient-weighted Class Activation Mapping) method was used by adopting an Explainable AI approach [17].

Fig.8 provides Grad-CAM-based visual evidence of how the proposed RDE-PneuNet model makes its predictions.

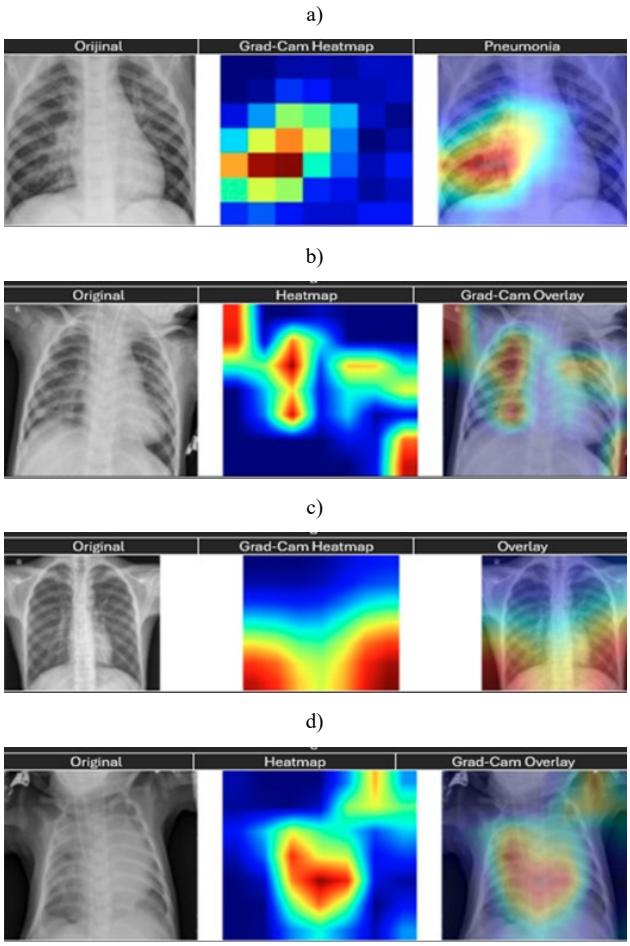


Fig. 8. Grad-CAM visualizations of the proposed RDE-PneuNet model. a) Correct localization on infected lung regions in a pneumonia case. b) Partial attention outside lung boundaries despite correct classification. c) Low activation in lung regions for a normal case. d) Strong activation highlighting severe pneumonia patterns.

As seen in Fig.8.a, the model accurately localizes the infected lung region in a pneumonia case. In Fig.8.b, the model still produces the correct prediction, although the attention map partially extends outside the lung boundaries, revealing a slight localization deviation. In contrast, Fig.8.c shows a normal chest X-ray with relatively low activation over the lung region, indicating that the model does not focus strongly on healthy tissue. Moreover, Fig.8.d demonstrates a severe pneumonia case, where the activation is concentrated over the pathological area, further supporting the clinical relevance and interpretability of the proposed model.

With Grad-CAM, class activation maps were generated by combining feature maps from the final convolutional layer with class-specific gradient information. By overlaying these maps onto the input images, it was visually confirmed that the model primarily focuses on lung regions and anatomically meaningful features. The analysis proved that the model has an interpretable and reliable structure, strengthening its potential for clinical use. Fig.8 is used to visualize Grad-CAM results for samples a, b, c, and d. The highlighted regions indicate the areas that contribute most to the model's decisions. The model consistently attends to relevant pulmonary regions, thereby validating its interpretability and strict alignment with clinical diagnostic expectations. Grad-CAM analysis reveals that the proposed RDE-PneuNet model consistently attends to clinically meaningful pulmonary regions in pneumonia cases while suppressing irrelevant activations in normal samples, thereby demonstrating strong discriminative capability and clinically interpretable decision-making, despite minor localization deviations

V. Discussion

Experimental evaluation confirms that the RDE-PneuNet ensemble establishes superior clinical equilibrium over individual baseline networks. While standalone models tend to over-detect pneumonia—yielding elevated false-positive rates—the proposed weighted soft-voting strategy significantly enhances healthy subject identification, securing a 92% specificity. This diagnostic balance minimizes patient anxiety and overtreatment while maintaining a high sensitivity of 93% to ensure critical pathologies are not overlooked. Unlike alternative frameworks that sacrifice specificity for raw accuracy, RDE-PneuNet delivers a clinically safer solution via an optimally stratified dataset without relying on overly complex hybrid topologies. Furthermore, the integration of Grad-CAM heatmaps resolves the inherent "black-box" limitations of standard CNNs by verifying that predictions are driven by anatomically relevant features, such as pulmonary opacities, rather than background artifacts.

Future research will focus on validating this framework using real-world clinical data from Kocaeli City Hospital under institutional ethics committee approval. To preserve patient confidentiality while assessing generalizability, cross-dataset validation and federated learning paradigms will be explored. Additionally, future iterations will investigate Vision Transformers (ViTs), lightweight deep learning architectures, and meta-heuristic optimization algorithms to refine the ensemble selection strategy. Ultimately, the objective is to optimize the computational footprint of RDE-PneuNet for embedded edge hardware deployment, transforming it into a portable, real-time clinical decision support system.

References

[1] B. Tripathy, S. Khan, S. Bebortta, A. Kamal, S. S. Tripathy, M. Fazil, and A. M. Albarrak, "TL-PneuNet: A transfer learning-based pneumonia classification framework," Scientific Reports, vol. 15, no. 1, pp. 1–12, 2025.

[2] A. Manickam, J. Jiang, Y. Zhou, A. Sagar, R. Soundrapandiyan, and R. D. J. Samuel, "Automated pneumonia detection on chest X-ray images: A deep learning approach with different optimizers and transfer learning architectures," Measurement, vol. 184, p. 109953, 2021.

[3] L. de F. A. Trindade, L. A. da Silva, M. L. C. S. Serrano, I. G. S. de A. Araújo, F. S. F. Prof., and V. C. Amador, "Comparative analysis of machine learning techniques for pneumonia detection in chest X-ray images," Cureus, vol. 17, no. 1, pp. 1–10, 2025.

[4] D. Rai, H. Verma, M. Arham, and S. Shukla, "A comprehensive survey on comparative analysis of machine learning algorithms in pneumonia detection," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 2, pp. 1–10, 2023.

[5] Y. S. Chowdhury, R. Dasgupta, and S. Nanda, "Analysis of various optimizers on CNN models in the application of pneumonia detection," Procedia Comput. Sci., vol. 218, pp. 123–130, 2022.

[6] R. Jain, P. Nagrath, G. Kataria, V. S. Kaushik, and D. J. Hemanth, "Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning," Measurement, vol. 165, p. 108046, 202The primarytt

[7] M. Aledhari, S. Joji, M. Hefeida, and F. Saeed, "Optimized CNN-based diagnosis system to detect pneumonia from chest radiographs," IEEE Access, vol. 8, pp. 1–10, 2020.

[8] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 4510–4520.

[9] H. Slimi, A. Balti, S. Abid, and M. Sayadi, "Trustworthy pneumonia detection in chest X-ray imaging through attention-guided deep learning," Scientific Reports, vol. 15, no. 1, pp. 1–15, 2025.

[10] S. Singh, M. Kumar, A. Kumar, B. K. Verma, and S. Shitharth, "Pneumonia detection with QCSA network on chest X-ray," Scientific Reports, vol. 13, no. 1, pp. 1–12, 2023.

[11] Z. Özdemir and H. Y. Keleş, "Derin öğrenme ile göğüs röntgeni görüntülerinde COVID-19 algılama," in Proc. 29th Signal Processing and Communications Applications Conf. (SIU), 2021, pp. 1–4.

[12] L. Zou *et al.*, "Ensemble Image Explainable AI (XAI) Algorithm for Severe Community-Acquired Pneumonia and COVID-19 Respiratory Infections," *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 2, pp. 242-254, April 2023, doi: 10.1109/TAI.2022.3153754.

[13] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell, vol. 172, no. 5, pp. 1122–1131, 2018.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. Learn. Representations (ICLR), 2015.

[15] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 4700–4708

[16] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Machine Learning (ICML), 2019, pp. 6105–6114.

[17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 618–626.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.

[19] P. Mooney, "Chest X-Ray Images (Pneumonia)," Kaggle, 2018. [Online]Available:https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

[20] Z.-H. Zhou, Ensemble Methods: Foundations and Algorithms, 2nd ed. Boca Raton, FL, USA: CRC Press, 2025.

Personel Kaynaklı Siber Risklerin Analizi ve Yapay Zekâ Destekli Siber Farkındalık Eğitim Modeli

1st Sinan Özçelik
Siber Güvenlik
Ankara Medipol Üniversitesi
Ankara, Türkiye
sinan.ozcelik@std.ankaramedipol.edu.tr (0009-0009-8059-7102)

2nd Cevat Özarpa
Siber Güvenlik
Ankara Medipol Üniversitesi
Ankara, Türkiye
cevat.ozarpa@ankaramedipol.edu.tr
(0000-0002-1195-2344)

***Özet* — Siber güvenlik ihlallerinde insan faktörü; kullanıcı hataları, farkındalık eksikliği ve sosyal mühendislik saldırılarına açıklık gibi unsurlar üzerinden kritik bir risk oluşturmaktadır. Mevcut yaklaşımlar çoğunlukla teknik önlemlerle sınırlı kalırken, kullanıcı davranışlarını sürekli analiz eden ve bu verilere göre dinamik eğitim üreten bütünleşik modellerin sınırlı olduğu görülmektedir. Bu çalışmada, söz konusu boşluğu gidermek amacıyla yapay zekâ destekli, davranış odaklı ve adaptif bir siber farkındalık eğitim modeli geliştirilmiş ve test edilmiştir. Bu çalışmanın temel katkısı, davranışsal risk profillemesi ile yapay zekâ tarafından oluşturulan kimlik avı simülasyonlarının aynı yapı içerisinde bütünleştirilmesidir. Ayrıca, gerçek zamanlı mikro-öğrenme geri bildirimi ile bu bileşenlerin kapalı döngülü bir eğitim modeli haline getirilmesi sağlanmıştır. Önerilen model; davranışsal veri toplama, risk profilleme, dinamik senaryo üretimi ve mikro-öğrenme temelli geri bildirim mekanizmalarını bütünleşik bir yapıda sunmaktadır. Modelin etkinliği kapsamında CERT veri seti üzerinde gerçekleştirilen simülasyon ile değerlendirilmiş; bu süreçte toplam 6649 olay işlenmiş ve risk temelli tetikleme mekanizması ile 29 oltalama senaryosu üretilmiştir. Kullanıcı tepkileri incelendiğinde, 10 olayda raporlama, 9 olayda yok sayma, 6 olayda bağlantıya tıklama ve 4 olayda zararlı ek indirme davranışı gözlemlenmiştir. Bulgular, kullanıcı davranışlarının büyük ölçüde kazara ve ihmalkâr eylemlerden kaynaklandığını göstermektedir. Elde edilen sonuçlar, modelin kullanıcı davranışlarını çok boyutlu analiz edebildiğini, risk skorlarını dinamik olarak güncelleyebildiğini ve eğitim müdahalelerini davranış temelli olarak tetikleyebildiğini ortaya koymaktadır. Bu yönüyle çalışma, statik farkındalık eğitimlerine alternatif olarak sürekli öğrenen, bağlama duyarlı ve kişiselleştirilmiş bir eğitim yaklaşımı sunmaktadır.**

Anahtar Kelimeler— siber güvenlik, siber farkındalık, siber farkındalık eğitimi, yapay zekâ, eğitim modeli, siber risk analizi

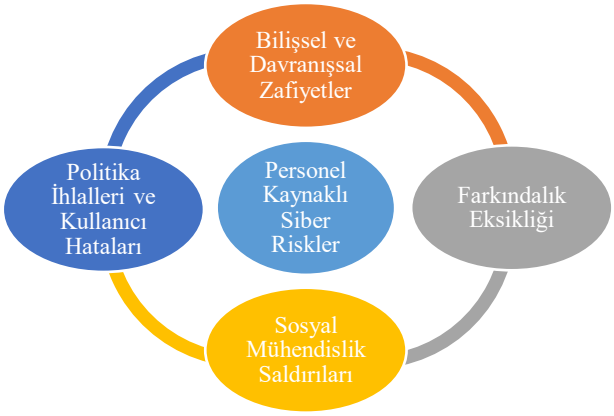
***Abstract* — The human factor in cybersecurity breaches poses a critical risk through elements such as user errors, lack of awareness, and vulnerability to social engineering attacks. While current approaches are mostly limited to technical measures, it is observed that integrated models which continuously analyze user behaviors and generate dynamic training based on these data are limited. In this study, to address this gap, an AI-supported, behavior-oriented, and adaptive cyber awareness training model has been developed and tested. The main contribution of this study is the integration of behavioral risk profiling and AI-generated phishing simulations within the same framework. Additionally, through real-time micro-learning feedback, these components have been transformed into a closed-loop training model. The proposed model offers behavioral data collection, risk profiling, dynamic scenario generation, and micro-learning-based feedback mechanisms in an integrated structure. Within the scope of the model's effectiveness, it was evaluated via a simulation conducted on the CERT dataset ; during this process, a total of 6649 events were processed, and 29 phishing scenarios were generated with a risk-based triggering mechanism. When user responses were examined, reporting in 10 events, ignoring in 9 events, clicking on links in 6 events, and downloading malicious attachments in 4 events were observed. The findings indicate that user behaviors largely stem from accidental and negligent actions. The obtained results reveal that the model can analyze user behaviors multi-dimensionally, update risk scores dynamically, and trigger training interventions on a behavioral basis. In this aspect, the study presents a continuously learning, context-sensitive, and personalized training approach as an alternative to static awareness trainings.**

Keywords — cybersecurity, cyber awareness, cyber awareness training, artificial intelligence, training model, cyber risk analysis

I. GİRİŞ

Günümüzün hızla dijitalleşen kurumsal ekosistemlerinde, bilgi sistemleri altyapılarının karmaşıklığı ve veri odaklı operasyonların hacmi benzeri görülmemiş bir düzeye ulaşmıştır. Kurumların gelişmiş ağ güvenliği sistemleri, şifreleme protokolleri ve saldırı tespit mekanizmalarına yaptıkları büyük ölçekli yatırımlara rağmen, bilgi güvenliği zincirinin en zayıf halkası olan insan faktörü, siber tehditlerin merkezinde yer almaktadır [1]. İnsan faktörü, siber güvenlikte "en zayıf halka" olarak görülmektedir [2]. Personel kaynaklı siber riskler; yalnızca basit kullanıcı hataları değil, aynı zamanda farkındalık eksikliği, güvenlik politikalarına uyumsuzluk ve sosyal mühendislik saldırılarına karşı savunmasızlık gibi çok katmanlı zafiyetlerden oluşmaktadır [3].

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE



Şekil 1. Personel kaynaklı siber risk yapısı

Şekil 1, personel kaynaklı siber risklerin temel bileşenlerini göstermektedir. Bu riskler; bilişsel ve davranışsal zafiyetler, farkındalık eksikliği, sosyal mühendislik saldırıları ile politika ihlalleri ve kullanıcı hatalarından oluşmaktadır.

Araştırmalar, şirketlerde meydana gelen bilgi güvenliği ihlallerinin önemli bir bölümünün doğrudan veya dolaylı olarak insan faktöründen kaynaklandığını göstermektedir [4]. Özellikle sosyal mühendislik saldırılarında, saldırganların bireylerin psikolojik eğilimlerini hedef alarak sistemlere erişim sağladığı ve bu tür saldırıların çoğu siber olayın başlangıç noktası olduğu vurgulanmaktadır [5]. İnsan davranışı ve farkındalık düzeyi, modern siber güvenlik yaklaşımlarının temel belirleyicilerinden biri haline gelmiştir.

Siber farkındalık eksikliğinin kurumlar üzerindeki finansal etkileri, geniş çaplı sektörel araştırmalar ve sayısal verilerle açık bir şekilde ortaya konmuştur. Yapılan analizler de büyük ölçekli bir şirkete yönelik gerçekleştirilen başarılı bir siber saldırının ortalama maliyetinin 2,6 milyon dolar seviyesinde olduğu görülmektedir. Ayrıca kurumların uğradığı veri ihlalinin toplam maliyeti ise yaklaşık 3,9 milyon dolara ulaştığını göstermektedir [6].

Personellerin siber farkındalık eksikliklerinin kurumlar üzerindeki finansal etkisi oldukça yüksektir. Siber farkındalıktan dolayı ortaya çıkan veri ihlallerinin ortalama maliyetinin 3,86 milyon ABD doları seviyesinde olduğu raporlanmıştır [7]. Bununla birlikte, küresel siber suç maliyetlerinin 2025 yılı itibarıyla 10,5 trilyon dolara ulaştığı görülmektedir [8].

Ek olarak, Ponemon Institute tarafından yürütülen ve Mart 2023 – Şubat 2024 döneminde veri ihlali yaşayan 604 farklı kurumu kapsayan analizler, veri ihlallerinde yalnızca dış kaynaklı saldırıların değil, aynı zamanda kurum içi personel hatalarının da önemli ölçüde maliyet yükü oluşturduğunu göstermektedir [9].

Bu kapsamda literatürde yer alan sayısal veriler Tablo 1 üzerinde özetlenmiştir.

Tablo 1. Siber güvenlik ihlallerine ilişkin sayısal bulgular

Gösterge	Değer	Açıklama	Kaynak
Ortalama veri ihlali maliyeti	3,86 milyon $	Kurum başına ortalama maliyet	[7]
Küresel siber suç maliyeti	10,5 trilyon $	2025 küresel etki	[8]
İnsan kaynaklı ihlal oranı	%77	İnsan zafiyetlerinden kaynaklı ihlaller	[4]
İnsan hatasına bağlı ihlaller	%98'e kadar	İnsan hatasının baskın rolü	[10]
Sağlık sektöründe insan hatası	%60–70	Kritik sektörlerde etki	[11]
Sosyal mühendislik ihlal maliyeti	4,10 milyon $	Ortalama ihlal maliyeti	[12]

Tablo 1 incelendiğinde, siber güvenlik ihlallerinin büyük ölçüde insan faktöründen kaynaklandığı ve bu ihlallerin kurumlar üzerinde ciddi finansal yükler oluşturduğu açıkça görülmektedir. Buna rağmen, şirketlerin siber güvenlik bütçelerinin önemli bir kısmı teknik altyapılara ayırmaktadır. İnsan odaklı güvenlik yaklaşımları ise çoğu zaman geri planda kaldığı yapılan çalışmalarda yer almaktadır [13].

Sektörel bazda incelendiğinde, sağlık hizmetleri gibi kritik altyapı sistemlerinde insan faktörü kaynaklı zafiyetlerin sonuçları çok daha önemli boyutlara ulaşmaktadır. İstatistiksel risk değerlendirmesi çalışmaları, sağlık sektöründeki siber güvenlik ihlallerinin yüzde 60 ila yüzde 70'inin doğrudan insan hatası olduğu gösterilmektedir [11].

Bu hataların başlıca nedenleri arasında zayıf parola kullanımı, hassas verilerin yanlış işlenmesi, yetkisiz bilgi paylaşımı ve oltalama saldırılarına maruz kalma yer almaktadır. Bu tür ihlallerin etkileri yalnızca finansal kayıplarla sınırlı kalmayıp, kritik sağlık hizmetlerinin aksamasına ve hasta güvenliğinin tehlikeye girmesine yol açabilmektedir. Nitekim, İngiltere sağlık sistemini hedef alan WannaCry saldırısı sonucunda iptal edilen randevular ve durdurulan hasta kabulleri nedeniyle yaklaşık 5,9 milyon sterlinlik doğrudan finansal kayıp yaşanmıştır [8].

Teknik güvenlik çözümleri gelişmeye devam etse de insan kaynaklı siber riskler tamamen ortadan kaldırılamamaktadır. İnsan faktörünün etkisi, siber risklerin teknik boyutun yanı sıra davranışsal ve organizasyonel yönlerinin de bulunduğunu ortaya koymaktadır. Bu durum, siber güvenlik alanında insan davranışlarını merkeze alan yaklaşımların geliştirilmesini zorunlu hale getirmiştir. İnsan faktörüne dayalı siber risklerin temel bileşenleri Tablo 2 üzerinde özetlenmiştir.

Tablo 2. Personel kaynaklı siber risk bileşenleri

Risk Kategorisi	Açıklama	Örnek Davranışlar
Bilişsel zafiyetler	Kullanıcıların psikolojik ve bilişsel açıkları	Acele karar verme, aşırı güven

Davranışsal hatalar	Bilinçsiz veya dikkatsiz kullanıcı davranışları	Zayıf şifre kullanımı, link tıklama
Farkındalık eksikliği	Siber tehditler hakkında bilgi yetersizliği	Oltalama saldırısını fark edememe
Politika ihlalleri	Güvenlik prosedürlerine uyumsuzluk	Yetkisiz erişim, veri paylaşımı
İç tehditler	Kasıtlı veya kasıtsız içeriden riskler	Veri sızıntısı, sabotaj

Personel kaynaklı siber riskler, yalnızca teknik zafiyetlerden değil; bireylerin bilişsel özellikleri, davranış biçimleri ve kurumsal güvenlik politikalarına uyum düzeylerinden de etkilenmektedir. Literatürde insan faktörünün siber güvenlik ihlallerindeki belirleyici rolü sıklıkla vurgulanmakta olup, bu risklerin sistematik şekilde sınıflandırılması etkin risk yönetimi ve farkındalık eğitimlerinin tasarlanması açısından önem taşımaktadır. Bu nedenle, çalışmada personel kaynaklı siber risklerin temel bileşenlerini ortaya koymak ve önerilen yapay zekâ destekli eğitim modelinin hangi risk alanlarına müdahale ettiğini göstermek amacıyla Tablo 2 oluşturulmuştur. Tablo 2'de, kurumlarda en sık karşılaşılan personel kaynaklı siber risk kategorileri ile bu kategorilere ilişkin tipik kullanıcı davranışları özetlenmektedir.

Bu çalışma, personel kaynaklı siber riskleri kapsamlı bir şekilde analiz etmeyi amaçlamaktadır. Çalışmada, bu risklerin yalnızca teknik değil; aynı zamanda davranışsal ve organizasyonel boyutlar içerdiğini ortaya koymaktadır. Bu doğrultuda, yapay zekâ destekli, veri odaklı ve adaptif bir siber farkındalık eğitim modeli önerilmektedir. Önerilen model, kullanıcı davranışlarını analiz ederek bireysel risk seviyelerini belirlemekte ve kişiye özgü eğitim içerikleri üretmektedir. Bu yaklaşım, mevcut eğitim yöntemlerinin sınırlılıklarını aşmayı hedeflemektedir.

Bununla birlikte literatürde yer alan mevcut çalışmalar incelendiğinde, siber farkındalık eğitimlerinin çoğunlukla temel düzeyde eğitim üzerinden sunulduğu ve kullanıcı davranışlarında kalıcı bir değişim oluşturmakta yetersiz kaldığı görülmektedir. Bu bağlamda bu çalışma, davranış analizi, yapay zekâ destekli içerik üretimi ve anlık mikro-öğrenme mekanizmalarını tek bir kapalı döngü içerisinde birleştirerek literatürdeki bu ayrışmayı ortadan kaldırmayı hedeflemektedir. Önerilen model, kullanıcı davranışlarından elde edilen verileri anlık olarak analiz etmekte; risk temelli tetiklenen eğitim içerikleri üretmekte ve bu içerikleri kullanıcı etkileşimi ile sürekli güncellenen adaptif bir öğrenme sürecine dönüştürmektedir. Bu yönüyle çalışma, yalnızca tehditleri tespit eden veya yalnızca eğitim sunan yaklaşımların ötesine geçerek, tespit ve öğrenme süreçlerini bütünleşik ve sürdürülebilir bir sistem mimarisi altında bir araya getirmektedir.

II. LİTERATÜR İNCELEMESİ

Siber güvenlik alanında son yıllarda öne çıkan en kritik yaklaşımlardan biri, siber farkındalık eğitim modellerinin geliştirilmesi ve bu modellerin davranışsal temelli olarak yeniden yapılandırılmasıdır. Literatürde, eğitim yaklaşımlarının kullanıcı davranışlarını kalıcı olarak değiştirmede yetersiz kaldığı; buna karşılık, simülasyon tabanlı, etkileşimli ve adaptif eğitim modellerinin kullanıcıların saldırılara karşı direnç seviyesini anlamlı ölçüde artırdığı açıkça ortaya konmuştur [14]. Özellikle oltalama saldırı simülasyonları ile desteklenen sürekli eğitim programlarının, çalışanların saldırılara karşı duyarlılığını önemli ölçüde azalttığı ve davranışsal öğrenmeyi güçlendirdiği belirtilmektedir [15]. Bununla birlikte, insan davranışlarına dayalı eğitim yaklaşımlarının bilişsel ve psikolojik faktörleri dikkate alarak tasarlanmasının, kullanıcı hatalarını azaltmada kritik rol oynadığı ve güvenlik farkındalığını sürdürülebilir hale getirdiği vurgulanmaktadır [10]. Ayrıca, sosyal mühendislik saldırılarına karşı geliştirilen farkındalık eğitim programlarının etkinliğinin sürdürülebilirliği için sürekli ve hedefe yönelik eğitim yaklaşımlarının gerekli olduğu vurgulanmaktadır [16]. Bu bağlamda, siber farkındalık eğitim modellerinin yalnızca bilgi aktarımı sağlayan araçlar değil, aynı zamanda çalışan davranışlarını analiz eden, öğrenme süreçlerini izleyen ve organizasyonel risk seviyesine göre sürekli güncellenen dinamik sistemler olarak ele alınması gerektiği anlaşılmaktadır.

A. Siber Risk

Siber riskler, teknolojik açıkların ötesinde insan davranışlarının yarattığı tehdit yüzeyi olarak değerlendirilmektedir. Siber saldırganlar, teknik zafiyetlerden ziyade insan hatası, güven duygusu ve yanlış bilgi gibi faktörleri sömürmekte; bu durum da kullanıcı davranış dinamiklerini ve karar alma süreçlerini içeren salgın tabanlı analitik modellemelerin kullanılmasını zorunlu kılmaktadır [17]. Sistem ne kadar iyi korunursa korunsun, kullanıcının oltalama saldırılarına yatkınlığı siber saldırı riskinin temel problemidir.

B. Siber Farkındalık

Kurumların olası saldırılardan korunmasında, personelin riskleri idrak edebileceği seviyeye gelmesi hayati önem taşımaktadır. Oltalama saldırıları, zayıf şifre tercihleri ve sistem açıklarının sömürülmesi gibi kimlik doğrulama ihlallerinin önlenebilmesi için kullanıcıların düzenli siber farkındalık eğitimleri alması gerektiği vurgulanmaktadır [18]. Ayrıca bilgi sistemleri altyapılarının da bu farkındalık anlayışıyla uyumlu şekilde çalışması gerektiği belirtilmektedir.

C. Yapay Zekâ ve Siber Güvenlik

Yapay zekâ sistemleri, tehditleri algılama ve önleme konusunda devrimsel fırsatlar sunmaktadır. Makine öğrenimi algoritmaları, siber güvenlik bağlamında anormal davranışları tespit etme, riskleri analiz etme ve tehditleri belirleme süreçlerinde önemli bir rol oynamaktadır [19]. Algoritmalar, güvenlik duvarlarının aşılmasını engellemek için hızlı ve otonom refleksler geliştirir.

Siber güvenlik eğitiminde ise yapay zekâ temelli yenilikler henüz erken aşamadadır. Örneğin, Żywiołek ve ark. (2025) tarafından Avrupa'da yapılan geniş ölçekli bir anket çalışmasında kullanıcılar, yapay zekâ destekli güvenlik eğitimlerinin değerli bulunduğunu, ancak hâlihazırda

"mecburi" veya "temel" beklentiler arasında görülmediğini belirtmişlerdir [20]. Bununla birlikte, kişiselleştirilmiş öğrenme ve adaptif eğitim mekanizmaları yapay zekâ ile sağlanabilmektedir. Örneğin Al-Kuwari (2024) çalışmasında eğitim ortamına gömülü bir ConvLSTM tabanlı model kullanılarak kullanıcı etkileşimleri analiz edilmiş, öğrenme süreci izlenmiş ve katılımcılara uyarlamalı geri bildirim sağlanmıştır. [21]. Bu tür yöntemler, katılımcı ihtiyaçlarına göre içeriği dinamik olarak optimize etmekte ve öğrenme sürecini güçlendirmektedir. Literatürde yapay zekâ destekli eğitim modelleri henüz sınırlı sayıda incelenmiştir; mevcut çalışmalarda genellikle ya insan faktörü ya da teknik unsurlar ele alınmıştır. Yapay zekâ destekli bir farkındalık modeli alanında kapsamlı araştırma eksikliği göze çarpmaktadır. Prümmer ve ark. (2024) da mevcut programların davranış değişikliği yaratmada yetersiz kaldığını rapor etmiş; bu eksikliği gidermek için yeni ve inovatif eğitim çözümlerinin gerekli olduğunu vurgulamıştır [22].

D. Personel Kaynaklı Siber Riskler

Bireylerin kurumsal ağ içerisindeki hataları en yaygın tehdit türüdür. Yapılan araştırmalar; siber güvenlik konularındaki temel bilgi eksikliğinin, risk alma eğilimlerinin, yorgunluğun, anlık duygusal durumların ve yaşa bağlı zafiyetlerin siber güvenlik ihlallerine yol açan en belirgin personel kaynaklı faktörler olduğunu kesin olarak doğrulamaktadır [23]. İnsan faktörünün yarattığı bu boşluklar, sadece yazılım güncellemeleriyle kapatılamamaktadır.

E. Siber Farkındalık Eğitimi Yaklaşımları

Mevcut eğitim yaklaşımları genellikle yüzeysel ve sıkıcı bulunmakta, çalışanda beklenen davranışsal değişimi yaratamamaktadır [24]. Literatürde, üniversiteler ve kurumsal yapılarda bilgi güvenliği ile kişisel verilerin korunmasına yönelik eğitimlerin yetersiz kaldığı belirtilmektedir. Siber savunmanın güçlendirilmesi için politika, bilinç ve eğitim düzeylerinin çok boyutlu çalışmalarla artırılması gerektiği vurgulanmaktadır.

F. Eğitim Modeli

Eğitim materyallerinin formatı ve oluşturulma biçimi, personelin konuyu içselleştirmesini doğrudan etkilemektedir. Bibliyometrik analizler, yapay zekânın siber savunma mekanizmalarında yaygın olarak kullanıldığını göstermektedir. Buna karşın, siber güvenlik eğitim programlarını geliştirme ve farklı bilgi seviyelerine sahip kullanıcılara uyarlanmış öğrenme deneyimleri sunma potansiyelinin yeterince değerlendirilmediği belirtilmektedir [25].

Literatürde yer alan çalışmalar, siber güvenlik alanında insan faktörünün önemli bir risk unsuru olduğunu göstermektedir. Özellikle kullanıcı davranışları, farkındalık eksikliği ve politika ihlalleri, kurumların güvenlik seviyesini doğrudan etkilemektedir. Bu nedenle modern yaklaşımlar, yalnızca teknik savunma mekanizmalarına değil, insan odaklı güvenlik modellerine de odaklanmaktadır.

Bununla birlikte mevcut çalışmaların önemli bir kısmı eğitim yöntemleri üzerine kuruludur. Kullanıcı davranışlarını sürekli izleyen ve bu verilere göre dinamik eğitim üreten sistemlerin sınırlı olduğu görülmektedir. Özellikle yapay zekâ destekli ve davranış temelli eğitim modelleri literatürde yeterince ele alınmamaktadır.

Bu eksiklik doğrultusunda önerilen model, davranış analizi, risk profilleme ve dinamik senaryo üretimini bütünleşik bir yapı içerisinde ele almaktadır. Böylece kullanıcıya özel, uyarlanabilir ve sürekli öğrenen bir siber farkındalık yaklaşımı sunulmaktadır.

Tablo 3, literatür taramasında tespit edilen mevcut durum ve tespit edilen eksiklikleri alt başlıklar özelinde özetlemektedir.

Tablo 3. Siber güvenlik eğitiminde mevcut durum ve literatür eksiklikleri

Alt Başlık	**İncelenen Konu / Mevcut Durum**	**Tespit Edilen Literatür Eksikliği**
Siber Risk	Tehdit yüzeyinin insan davranışları üzerinden tanımlanması.	Bilişsel açıkların teknik risk analizlerine yeterince dâhil edilmemesi.
Siber Farkındalık	Kurum içi farkındalık kampanyalarının önemi.	Bilgi aktarımının personelin iş akışına anlık olarak entegre edilememesi.
Yapay Zekâ ve Siber Güvenlik	Sıfırıncı gün zafiyetlerinde otonom makine öğrenimi kullanımı.	Yapay zekânın ağırlıklı olarak ağ savunmasında kalıp insanı merkeze almaması.
Personel Kaynaklı Siber Riskler	Yorgunluk, risk alma eğilimi ve stresin hataları tetiklemesi.	Psikolojik faktörleri algoritmik bir skorlama sistemine dönüştüren yapıların azlığı.
Siber Farkındalık Eğitimi Yaklaşımları	Tekdüze seminerler ve periyodik genel testler.	Çalışanlarda "güvenlik yorgunluğu" yaratan sabit materyallerin etkisizliği.
Eğitim Modeli	Kurumsal uyumluluk ve denetim odaklı eğitim süreçleri.	Yapay zekânın siber güvenlik eğitim programlarında etkin ve yaygın biçimde kullanılmaması.

Literatürde yer alan çalışmalar, siber güvenlik alanında insan faktörü, farkındalık eğitimleri ve yapay zekâ uygulamaları üzerine önemli katkılar sunmaktadır. Ancak mevcut yaklaşımların büyük ölçüde teknik savunma mekanizmalarına veya genel farkındalık eğitimlerine odaklandığı görülmektedir. Buna karşın, kullanıcı davranışlarını sürekli izleyen, risk seviyelerini dinamik olarak değerlendiren ve bu değerlendirmelere göre kişiselleştirilmiş eğitim içerikleri üreten bütünleşik sistemler literatürde sınırlı düzeyde yer almaktadır. Bu nedenle Tablo 3, mevcut çalışmaların odaklandığı konuları ve bu çalışmalar

kapsamında belirlenen temel eksiklikleri karşılaştırmalı olarak ortaya koymak amacıyla hazırlanmıştır. Tablo incelendiğinde, özellikle davranış analizi, risk profilleme ve yapay zekâ destekli dinamik eğitim üretimi bileşenlerinin bütünleşik bir yapıda ele alınmadığı görülmektedir. Bu eksiklikten hareketle, bu çalışmada davranışsal veri analizi ile yapay zekâ destekli içerik üretimini birleştiren yeni bir siber farkındalık eğitim modeli önerilmektedir.

III. METOT VE MATERYAL

Bu çalışmada ortaya konulan araştırma sorularını yanıtlamak amacıyla, personelin risk seviyesini gerçek zamanlı izleyen ve eğitim materyalini bu verilere göre dinamik biçimde şekillendiren "Yapay Zekâ Destekli Yeni Siber Farkındalık Eğitim Modeli" tasarlanmıştır.

Bu çalışmada kullanılan davranışsal veri kümesi, Carnegie Mellon University Software Engineering Institute tarafından geliştirilen CERT Insider Threat Dataset (CERT veri seti) kullanılmıştır [26]. Bu veri seti, kullanıcı logon kayıtları, e-posta etkileşimleri, cihaz kullanımı, dosya erişimleri ve web aktiviteleri gibi çok boyutlu davranışsal verileri içermektedir. Veri setinin tercih edilmesinin temel nedeni, gerçek kurum içi kullanıcı davranışlarını yüksek doğrulukla simüle edebilmesi ve iç tehdit analizleri için zengin bir davranışsal iz sunmasıdır.

CERT veri seti, gerçek dünya iç tehdit olaylarından türetilmiş zengin davranışsal veriler sunarak, bu tür tehditlerin analiz edilmesi ve model geliştirme süreçlerinde yaygın olarak kullanılan önemli bir araştırma kaynağıdır [27]. Bu kapsamda CERT veri seti, önerilen modelin dinamik risk değerlendirme ve adaptif eğitim mekanizmalarının test edilmesi için uygun ve güvenilir bir deneysel ortam sağlamaktadır.

A. Eğitim Modelinde Kullanılan Temel Kavramlar

Bu bölümde, tasarlanan modelin algoritmik ve eğitimsel temelini oluşturan, literatürde geçerliliği kanıtlanmış kavramlar açıklanmaktadır.

1) K-Means

K-Means algoritması, veri setlerini özellik benzerliklerine göre belirli sayıda kümeye ayırmayı amaçlayan denetimsiz bir makine öğrenimi yöntemidir. Algoritma, benzer özelliklere sahip veri noktalarını aynı kümede toplarken farklı özelliklere sahip verileri ayrı kümelere atamaktadır [28].

2) Mean Shift

Mean Shift, veri noktalarının yoğunluk fonksiyonunun gradyanını kullanarak yerel maksimum noktalarına doğru iteratif bir şekilde hareket etmesine dayanan parametrik olmayan bir kümeleme algoritmasıdır. Algoritma, her bir veri noktası için bir çekirdek fonksiyonu aracılığıyla komşuluk bölgesini tanımlar ve bu bölge içindeki noktaların ağırlıklı ortalamasını bularak noktayı yoğunluğun en yüksek olduğu bölgeye doğru kaydırır. Bu süreç, noktalar yoğunluk yüzeyinin tepe noktalarına ulaşana kadar tekrarlanır [29].

3) Eğitim Üretim Motoru

Kullanıcının bilgi düzeyine, rolüne ve davranışsal risk skoruna göre dinamik olarak eğitim materyali üreten yapay zekâ tabanlı bir sistem bileşenidir. Büyük dil modelleri, siber farkındalık eğitimlerini desteklemek amacıyla gerçekçi ve detaylı sosyal mühendislik senaryoları üreterek uygulamalı öğrenme süreçlerine katkı sağlamaktadır [30].

4) Naive Bayes Algoritması

Bayes teoremine dayanan ve özellikler arasında bağımsızlık varsayımı yapan olasılıksal bir sınıflandırma algoritmasıdır. Siber güvenlik ortamlarında, anomali tespiti ve ağ saldırılarının sınıflandırılması gibi problemlerde hızlı, esnek ve verimli sonuçlar üretmektedir [31].

5) Doğal Dil İşleme

Doğal Dil İşleme, yapay zekânın insan dilini anlaması, yorumlaması ve işlemesini sağlayan bir alan olup, siber güvenlikte özellikle oltalama e-postalarının tespit edilmesinde yaygın olarak kullanılmaktadır. Yapılan çalışmalar, TF-IDF, Word2Vec ve BERT gibi metin temsili yöntemleri ile Naive Bayes, SVM, LSTM ve ELM gibi sınıflandırma algoritmalarının bir arada kullanılmasının, oltalama saldırısı tespit başarımını anlamlı ölçüde artırdığını göstermektedir. Özellikle eklemeli yapısı nedeniyle zengin morfolojik çeşitlilik gösteren Türkçe gibi dillerde, bağlamsal dil modellerinin (BERTurk gibi) mevcut yöntemlere kıyasla daha yüksek doğruluk oranlarına ulaştığı deneysel olarak ortaya konmuştur [32].

6) Sosyal Mühendislik ve Oltalama Saldırısı

Saldırganların sistemlerdeki teknik zafiyetleri aramak yerine doğrudan insan psikolojisini, güven duygusunu ve zafiyetlerini manipüle ederek hassas bilgileri ele geçirmeyi amaçladığı e-posta veya mesaj odaklı siber saldırı türüdür [33].

7) Mitre Att&Ck

MITRE ATT&CK çerçevesi, siber saldırganların kullandığı taktikler, teknikler ve prosedürleri sistematik bir şekilde sınıflandıran kapsamlı bir bilgi tabanı olarak tanımlanmaktadır. Bu yapı, saldırgan davranışlarının analiz edilmesine olanak tanıyarak siber tehdit istihbaratı, saldırı tespiti, olay müdahalesi ve risk değerlendirme gibi çeşitli alanlarda yaygın olarak kullanılmaktadır. Ayrıca ATT&CK hem akademik çalışmalarda hem de endüstriyel uygulamalarda farklı kullanım senaryoları ile benimsenmiş olup, siber güvenlik araştırmalarında önemli bir referans çerçevesi sunmaktadır [34].

8) iCAT (Integrated Cybersecurity Awareness Training)

Oyunlaştırma dinamiklerini ve mikro-öğrenme prensiplerini kullanarak siber güvenlik eğitimlerini daha esnek, erişilebilir ve çalışan katılımı odaklı hale getiren bütünleşik bir eğitim yaklaşımıdır [35].

9) Bloom Taksonomisi

Eğitim hedeflerini hatırlama, anlama, uygulama, analiz etme, değerlendirme ve yaratma şeklinde basitten karmaşığa doğru bilişsel seviyelere ayıran, e-öğrenme süreçlerini yapılandıran uluslararası bir eğitim sınıflandırma sistemidir [36].

B. Eğitim Modelinin Çalışma Yapısı ve Detayları

Tanımlanan temel kavramlar üzerinden geliştirilen eğitim modeli üç ana aşamada çalışmakta ve personeli kesintisiz bir geri bildirim döngüsüne dâhil etmektedir. Bu modelin genel çalışma mantığı ve bileşenler arası etkileşim, Şekil 2 üzerinde

gösterilen kapalı döngü yapısı üzerinden kavramsallaştırılmıştır.



Şekil 2. Yapay zekâ destekli siber farkındalık eğitim modelinin kapalı döngü yapısı

Şekil 2 incelendiğinde, sistemin çalışan davranışlarından elde edilen verilerle başladığı görülmektedir. Bu veriler analiz edilerek risk skoruna dönüştürülmekte ve elde edilen risk profiline göre kullanıcıya özel saldırı simülasyonları üretilmektedir. Kullanıcının bu simülasyonlara verdiği tepki, sistem tarafından anlık olarak değerlendirilmekte ve uygun mikro-öğrenme içerikleri ile geri bildirim sağlanmaktadır. Bu sürecin sonunda kullanıcının davranışları güncellenmekte ve elde edilen yeni davranış verileri tekrar sisteme dâhil edilerek döngü devam etmektedir.

Bu yapı, önerilen modelin statik ve tek yönlü eğitim yaklaşımlarından farklı olarak sürekli öğrenme, adaptasyon ve davranış temelli iyileştirme sağlayan dinamik bir sistem olduğunu ortaya koymaktadır.

1) Birinci Aşama: Davranışsal Veri Toplama ve Risk Profilleme

İlk aşamada, personelin kurum içi sistemlerle etkileşimini yansıtan davranışsal veriler; ağ logları, kullanıcı aktiviteleri ve kimlik doğrulama kayıtları üzerinden anonimleştirilerek toplanmaktadır. Bu kapsamda; çalışanların sisteme erişim sıklıkları, hatalı parola girişleri, mesai dışı sistem kullanımları, erişim bağlamındaki anormallikler ve daha önce karşılaştıkları şüpheli durumlara verdikleri tepkiler bütüncül olarak analiz edilmektedir. Elde edilen bu çok boyutlu veri kümesi, kullanıcıların davranış örüntülerini ortaya çıkarmak ve bireysel risk seviyelerini belirlemek amacıyla istatistiksel ve davranışsal analiz yöntemleri ile işlenmektedir. Bu süreçte, her bir kullanıcının normal davranış profili oluşturularak, mevcut davranışları ile bu referans profil arasındaki sapmalar ölçülmekte ve bu sapmalar nicel bir risk göstergesine dönüştürülmektedir. Böylece bireylerin ihmalkâr, hataya açık veya potansiyel olarak zararlı davranışlar sergileme eğilimleri matematiksel bir risk modeline yansıtılmaktadır. Elde edilen risk skorları doğrultusunda personel; düşük, orta, yüksek ve kritik risk seviyelerine ayrılmakta ve bu sınıflandırma gerçek zamanlı olarak güncellenerek organizasyonun genel güvenlik durumu dinamik biçimde izlenmektedir.

Bu aşama, önerilen yapay zekâ destekli iç tehdit risk yönetimi modelinin temelini oluşturan veri toplama, davranışsal analiz ve risk profilleme süreçlerini bütüncül bir yapı içerisinde ele almaktadır. Amaç, kurum içerisindeki kullanıcıların dijital ortamdaki etkileşimlerini çok boyutlu olarak analiz ederek, her birey için dinamik, ölçülebilir ve sürekli güncellenen bir risk profili oluşturmaktır. Mevcut güvenlik yaklaşımlarının aksine, yalnızca teknik log verilerine dayanan analizlerin iç tehditleri tespit etmede yetersiz kaldığı; davranışsal, bağlamsal ve psikososyal göstergelerin de değerlendirmeye dahil edilmesi gerektiği literatürde açıkça ortaya konmaktadır [37]. Bu doğrultuda, sistem; kullanıcı davranışlarını temsil eden heterojen veri kaynaklarını entegre ederek, çok katmanlı bir risk değerlendirme yaklaşımı benimsemektedir.

Model kapsamında veri toplama sürecinde, sistem logları, kullanıcı aktiviteleri, kimlik doğrulama kayıtları, veri erişim hareketleri ve ağ etkileşimleri gibi farklı kaynaklardan veriler elde edilmektedir. Toplanan yüksek hacimli veriler merkezi bir veri havuzunda bir araya getirilmektedir. Bu veriler; zaman damgası hizalama, normalizasyon ve olay korelasyonu gibi ön işleme adımlarından geçirilerek analize hazır hale getirilmektedir. Böylece farklı sistemlerden gelen verilerin bütünleşik bir davranış modeli oluşturması sağlanmaktadır. Toplanan veriler yalnızca teknik aktiviteleri değil, aynı zamanda kullanıcının çalışma alışkanlıklarını, erişim davranışlarını ve sistem kullanım örüntülerini de yansıtan bir davranışsal iz oluşturmaktadır.

Davranışsal veri toplama sürecinin ardından sistem, her kullanıcı için bir "normal davranış profili" oluşturmaktadır. Bu profil, kullanıcının geçmiş aktivitelerine dayalı istatistiksel ve makine öğrenmesi temelli yöntemlerle geliştirilmekte ve kullanıcının tipik davranış dağılımını temsil etmektedir. Bu bağlamda kullanıcı davranışı bir özellik vektörü ile ifade edilmektedir:

$$X_u = \{x_1, x_2, x_3, \ldots, x_n\} \quad (1)$$

X_u , kullanıcıya ait davranışsal özellikler kümesini; $x_1, x_2, x_3, \ldots, x_n$ ise sırasıyla oturum süresi, erişim sıklığı, veri transfer miktarı, işlem türü, lokasyon ve cihaz özellikleri gibi davranışsal parametreleri temsil etmektedir. Bu özellikler zaman serisi olarak analiz edilmekte ve kullanıcının normal davranış dağılımı denklem 2 ile modellenmektedir:

$$P_u(x) = N(\mu_u, \sigma_u^2) \quad (2)$$

μ_u kullanıcının ortalama davranışını, σ_u ise davranış varyansını ifade etmektedir. Bu model, kullanıcının alışılmış davranış aralığını belirlemek için referans bir çerçeve sunmaktadır.

Risk profilleme sürecinin temelini, mevcut davranış ile normal davranış arasındaki sapmanın ölçülmesi oluşturmaktadır. Bu sapma, anomali skoru olarak tanımlanmakta ve tekil özellik bazlı değerlendirme denklem 3 ile hesaplanmaktadır:

$$A_u(t) = \frac{|x_t - \mu_u|}{\sigma_u} \quad (3)$$

$A_u(t)$, belirli bir zamanda gözlemlenen davranışın anomali derecesini ifade etmekte olup, x_t ilgili davranışın gözlemlenen değerini, μ_u kullanıcının normal davranış ortalamasını ve σ_u ise bu davranışın standart sapmasını temsil etmektedir. Yüksek anomali skorları, kullanıcının normal davranış örüntüsünden anlamlı şekilde saptığını ve potansiyel bir risk oluşturduğunu göstermektedir.

Ancak kullanıcı davranışları genellikle çok boyutlu bir yapı sergilediğinden, tekil özellik bazlı analizler tüm risk dinamiklerini yeterli düzeyde yansıtmayabilmektedir. Bu nedenle, birden fazla davranışsal özelliğin birlikte değerlendirildiği durumlarda birleşik anomali skoru denklem 4 ile tanımlanmaktadır:

$$A_u(t) = \sqrt{\sum_{i=1}^{n} \left(\frac{x_i(t) - \mu_i}{\sigma_i}\right)^2} \quad (4)$$

Denklemde, kullanıcının farklı davranışsal özelliklerinin birlikte değerlendirilmesini sağlayan çok boyutlu bir anomali ölçümünü temsil etmektedir. Burada $x_i(t)$, kullanıcının i. davranışsal özelliğinin $t$anındaki gözlemlenen değerini, μ_i ilgili özelliğin normal ortalamasını ve σ_i ise bu özelliğin standart sapmasını ifade etmektedir. n ise toplam davranışsal özellik sayısını gösterirken, i her bir özelliği temsil eden indekstir. Bu yapı sayesinde farklı ölçeklerdeki davranışlar normalize edilerek ortak bir ölçekte değerlendirilmekte ve kullanıcının genel davranış örüntüsündeki sapmalar bütüncül bir şekilde analiz edilebilmektedir. Böylece sistem, yalnızca tekil sapmaları değil, davranışların bütünsel değişimini dikkate alarak daha doğru ve güvenilir bir risk değerlendirmesi gerçekleştirebilmektedir.

Davranışsal analiz sonucunda elde edilen anomali skorları, bağlamsal ve organizasyonel faktörlerle birlikte değerlendirilerek birleşik bir risk skoru hesaplanmaktadır. Bu kapsamda önerilen genel risk fonksiyonu denklem 5 ile tanımlanmaktadır:

$$R_u = \sum_{i=1}^{n} w_i \cdot f_i(x_i) \quad (5)$$

R_u kullanıcıya ait toplam risk skorunu, w_i ilgili risk faktörünün ağırlığını ve $f_i(x_i)$ ise davranışsal veya bağlamsal risk fonksiyonunu ifade etmektedir. Bu yapı; kullanıcı yetki seviyesi, işlem türü, erişim bağlamı, zaman faktörü, cihaz güvenliği ve kümülatif davranış yoğunluğu gibi çoklu parametrelerin birlikte değerlendirilmesini sağlamaktadır.

Risk skorunun zamansal olarak daha anlamlı hale getirilebilmesi amacıyla modele dinamik bir güncelleme mekanizması da entegre edilmiştir. Bu doğrultuda kullanıcının risk seviyesi denklem 6 üstel ağırlıklı fonksiyon ile güncellenmektedir:

$$R_u(t) = \alpha \cdot R_u(t-1) + (1-\alpha) \cdot A_u(t) \quad (6)$$

Burada α geçmiş davranışların etkisini belirleyen öğrenme katsayısını temsil etmektedir. Bu yaklaşım sayesinde sistem, yalnızca anlık anomalilere değil, aynı zamanda davranışın sürekliliğine de duyarlı hale gelmektedir.

Elde edilen risk skorları, kullanıcıların önceliklendirilmesi ve güvenlik aksiyonlarının belirlenmesi amacıyla belirli eşik değerlerine göre sınıflandırılmaktadır:

- $0 \leq R_u < T_1 \rightarrow$ Düşük Risk
- $T_1 \leq R_u < T_2 \rightarrow$ Orta Risk
- $T_2 \leq R_u < T_3 \rightarrow$ Yüksek Risk
- $R_u \geq T_3 \rightarrow$ Kritik Risk

Bu eşik değerleri organizasyonun güvenlik politikalarına göre dinamik olarak belirlenmekte ve sistem tarafından sürekli güncellenmektedir. Böylece güvenlik ekipleri, yüksek riskli kullanıcıları hızlı bir şekilde tespit edebilmekte ve gerekli önleyici aksiyonları zamanında alabilmektedir.

Ayrıca sistem, kullanıcı davranışlarını niyet ve etki düzeyine göre sınıflandırmakta ve iç tehditleri kasıtlı, ihmalkâr ve kazara olmak üzere üç temel kategori altında incelemektedir. Bu yaklaşım, kullanıcı aktivitelerinin yalnızca teknik anomali düzeyinde değil, aynı zamanda davranışsal bağlam ve niyet perspektifinde de analiz edilmesini sağlamaktadır [38].

2) İkinci Aşama: Üretken Yapay Zekâ ile Dinamik Senaryo Üretimi

İkinci aşamada, belirlenen risk gruplarına özel olarak içerik sunacak "Eğitim Üretim Motoru" devreye girer. Kullanıcının yeterlilik veya risk profiline göre uyarlanan oltalama saldırısı eğitimlerinin etkinliğini inceleyen çalışmalar, kişiselleştirmenin eğitim tasarımında anlamlı bir iyileştirme alanı sunduğunu göstermektedir [39]. Kullanıcıların risk skorlarını sürekli güncelleyen Naive Bayes algoritması, kullanıcının o anki durumunu analiz ederek dinamik bir risk değerlendirmesi gerçekleştirir. Eğer kullanıcı "yüksek riskli" bir kümede yer alıyorsa, ona genel geçer bir eğitim dokümanı sunmak yerine, doğal dil işleme algoritmaları kullanılarak ilgili personelin departmanına, kullandığı mesleki dile ve günlük iş akışına tamamen uyumlu, yüksek gerçeklik düzeyine sahip sosyal mühendislik oltalama simülasyonları oluşturulur. Bu senaryolar, MITRE ATT&CK çerçevesinde tanımlanan güncel tehdit vektörlerini yansıtacak şekilde otonom olarak üretilmektedir. Bu yaklaşım, büyük dil modeli tabanlı oltalama farkındalık eğitimlerinin kullanıcıların oltalama saldırılarını tespit etme performansında ölçülebilir iyileşmeler sağladığını ortaya koyan deneysel bulgularla desteklenmektedir [40]. Literatürde yapılan çalışmalar, farklı içerik türleri ve senaryo yapılarının kullanıcıların oltalama saldırılarına verdiği tepkiler üzerinde doğrudan etkili olduğunu göstermektedir [41].

Üretilen bu senaryoların kullanıcıya ulaştırılması ve etkileşimlerin izlenmesi süreci, yalnızca sunucu taraflı bir içerik üretimi ile sınırlı kalmayıp, kullanıcı cihazları üzerinde çalışan bir istemci yazılımı ile desteklenmektedir. Bu istemci bileşeni, kullanıcının e-posta etkileşimleri, bağlantı tıklamaları ve sistem üzerindeki kritik davranışlarını arka planda sürekli olarak izleyerek, oluşturulan senaryolar ile gerçek kullanıcı davranışları arasında doğrudan bir bağ kurulmasını sağlamaktadır. Böylece eğitim süreci, kullanıcıdan bağımsız bir platformda değil, doğrudan kullanıcının günlük çalışma ortamına bütünleşmiş bir biçimde yürütülmektedir.

Sistem, üretilen oltalama senaryolarını kullanıcının mevcut iletişim kanalları üzerinden (kurumsal e-posta sistemleri, dahili mesajlaşma araçları veya web tabanlı bildirimler) ileterek simülasyonun gerçeklik düzeyini artırmaktadır. Kullanıcının bu senaryolarla etkileşimi, uç nokta izleme ajanı tarafından anlık olarak izlenmekte ve analiz edilmektedir. Özellikle kullanıcı tarafından gerçekleştirilen bağlantı tıklama, dosya indirme veya kimlik

bilgisi girme gibi riskli etkileşimler, sistem tarafından kritik olaylar olarak değerlendirilmektedir.

Bu yapı sayesinde eğitim üretim motoru yalnızca içerik üreten pasif bir bileşen olmaktan çıkmakta; kullanıcı davranışlarını gerçek zamanlı olarak izleyen ve bu davranışlara anında tepki verebilen aktif bir güvenlik mekanizmasına dönüşmektedir. Ayrıca kullanıcı etkileşimlerinden elde edilen veriler tekrar sisteme beslenerek, hem Naive Bayes tabanlı risk modelinin güncellenmesine hem de gelecekte üretilecek senaryoların daha hedefli ve etkili hale gelmesine katkı sağlamaktadır. Böylece sistem, statik içerik üretimi yerine sürekli öğrenen ve kullanıcıya uyum sağlayan dinamik bir eğitim ekosistemi oluşturmaktadır.

3) Üçüncü Aşama: Mikro-Öğrenme ve Etkileşimli Geri Bildirim

Son aşamada, kullanıcının sistem tarafından üretilen simülasyona (örneğin sahte bir oltalama e-postasına) verdiği tepki saniyeler içerisinde analiz edilir. Bu analiz süreci, kullanıcı cihazı üzerinde çalışan uç nokta izleme ajanı tarafından gerçek zamanlı olarak gerçekleştirilmekte ve kullanıcının gerçekleştirdiği etkileşimler (bağlantı tıklama, dosya indirme, form doldurma vb.) anlık olarak değerlendirmeye alınmaktadır.

Eğer kullanıcı zararlı olabilecek bir bağlantıya tıklar veya senaryo kapsamında riskli bir davranış sergilerse, sistem bu etkileşimi kritik bir güvenlik olayı olarak algılayarak anında müdahale eder. Bu müdahale sırasında, ilgili işlem güvenli bir şekilde durdurulur veya izole edilir ve kullanıcıya eş zamanlı olarak eğitim arayüzü sunulur. Böylece potansiyel bir güvenlik ihlali gerçekleşmeden önce süreç kontrollü bir eğitim deneyimine dönüştürülmektedir.

Bu noktada devreye giren mikro-öğrenme modülü, iCAT modelinin öğretimsel prensiplerine uygun olarak tasarlanmış etkileşimli bir eğitim katmanı sunmaktadır. Eğitim içeriği, kullanıcının yaptığı hatanın türüne, risk seviyesine ve bağlamsal faktörlere göre dinamik olarak belirlenmekte ve genellikle 30–60 saniye aralığında kısa, odaklı ve yüksek etki düzeyine sahip içeriklerden oluşmaktadır. Bu içerikler; kısa video anlatımları, etkileşimli senaryo geri bildirimleri, doğru davranışın gösterildiği mini simülasyonlar veya hızlı karar verme testleri şeklinde sunulabilmektedir.

Sunulan eğitim arayüzü, kullanıcının dikkatini dağıtmadan doğrudan çalışma ekranı üzerinde açılan bir katman veya kontrollü bir bildirim penceresi şeklinde tasarlanmıştır. Literatürde gömülü oltama saldırısı eğitimlerinin etkisinin özellikle anlık hatırlatma ve davranışsal dürtüleme mekanizmalarıyla güçlendiği gösterilmiştir [42]. Bu sayede kullanıcı, iş akışından tamamen kopmadan, yaptığı hatanın nedenini anlık olarak kavrayabilmekte ve doğru davranışı bağlam içerisinde öğrenmektedir. Benzer şekilde, büyük dil modeli tarafından üretilen açıklayıcı oltalama saldırısı uyarılarının kullanıcı kırılganlığını azaltmada etkili olabildiği büyük ölçekli kontrollü deneylerle gösterilmiştir [43]. Bloom Taksonomisi'nin üst düzey bilişsel basamakları dikkate alınarak hazırlanan bu mikro içerikler, kullanıcının yalnızca bilgiyi hatırlamasını değil; aynı zamanda analiz etmesini, değerlendirmesini ve doğru karar verme refleksi geliştirmesini hedeflemektedir.

Ayrıca sistem, kullanıcıya yalnızca pasif bir eğitim sunmakla kalmayıp, kısa doğrulama adımları ile öğrenme sürecini aktif hale getirmektedir. Bu yaklaşım, yapay zekâ destekli tehdit tespiti ile kullanıcıya yönelik eğitim mekanizmalarının entegre edildiği sistemlerin etkinliğini ortaya koyan güncel çalışmalarla örtüşmektedir [44]. Örneğin kullanıcıdan, gördüğü e-postadaki şüpheli unsurları işaretlemesi veya benzer bir senaryoda doğru davranışı seçmesi istenebilmektedir. Etkileşimli eğitim yaklaşımlarının yalnızca tehdit tanıma değil, aynı zamanda raporlama ve destek arama davranışlarını da güçlendirebildiği gösterilmiştir [45]. Bu etkileşimli yapı sayesinde öğrenme süreci pekiştirilmekte ve davranışsal farkındalık kalıcı hale getirilmektedir.

Gerçekleştirilen her kullanıcı etkileşimi ve eğitim sonucu, sistem tarafından kayıt altına alınarak davranışsal profil güncelleme sürecine geri besleme olarak aktarılmaktadır. Bu sayede kullanıcıların zaman içerisindeki gelişimi izlenebilmekte, eğitim etkinliği ölçülebilmekte ve sistemin gelecekte üreteceği senaryolar daha hedefli hale getirilebilmektedir. Böylece model, yalnızca anlık müdahale sağlayan bir yapı olmaktan çıkmakta; sürekli öğrenen ve kullanıcı davranışlarını iyileştirmeye odaklanan adaptif bir eğitim döngüsüne dönüşmektedir.

Tablo 4, tasarlanan bu eğitim modelinin temel aşamalarını ve bileşenlerini detaylandırmaktadır.

Tablo 4. Yapay zekâ destekli siber farkındalık eğitim modelinin aşamaları ve sistem bileşenleri

Eğitim modeli aşaması	**Kullanılan teknoloji / sistem bileşeni**	**Aşamanın amacı ve çıktısı**
Aşama 1: Davranışsal profilleme	Kümeleme algoritmaları, ortalama kaydırma yöntemi, davranış analizi modülü	Kullanıcı davranışlarını analiz ederek düşük, orta ve yüksek risk profilleri oluşturur
Aşama 2: Dinamik senaryo üretimi ve dağıtımı	Olasılıksal sınıflandırma yöntemi, doğal dil işleme, merkezi yapay zekâ motoru, uç nokta izleme ajanı	Risk skoruna göre kullanıcıya özel, departman bazlı ve gerçekçi oltalama senaryoları üretir ve kullanıcıya e-posta, mesaj veya sistem bildirimleri üzerinden iletir
Aşama 3: Gerçek zamanlı müdahale ve mikro öğrenme	Uç nokta izleme ajanı, davranış analizi modeli, mikro öğrenme arayüzü, etkileşimli eğitim modülü	Kullanıcının riskli etkileşimini anlık olarak algılar, işlemi

		durdurur ve kısa, etkileşimli eğitim içerikleri ile anında geri bildirim sağlar
Sürekli öğrenme ve uyum sağlama	Olasılıksal güncelleme yöntemi, davranışsal geri besleme, öğrenen model mekanizması	Kullanıcı tepkilerini analiz ederek risk skorlarını ve eğitim içeriklerini sürekli günceller, sistemi uyarlanabilir hale getirir
Standartlaştırma ve tehdit uyumu	Öğrenme hedefleri sınıflandırması, siber tehdit taktikleri ve teknikleri çerçevesi	Eğitim içeriğinin öğretimsel olarak etkili ve güncel tehdit modelleriyle uyumlu olmasını sağlar

Önerilen modelin aşamaları yalnızca teorik bir eğitim akışını değil, aynı zamanda gerçek zamanlı çalışan bir sistem mimarisini temsil etmektedir. Tablo 4 ile görüldüğü üzere, model; davranışsal veri analizi, yapay zekâ destekli içerik üretimi ve anlık müdahale mekanizmalarını bütünleşik bir yapı içerisinde ele almaktadır. Özellikle ikinci ve üçüncü aşamalarda yer alan uç nokta izleme ajanı bileşeni, sistemin kullanıcı cihazı üzerinde aktif olarak çalışmasını sağlayarak eğitim sürecini pasif bir içerik sunumundan çıkarıp, doğrudan kullanıcı davranışına bağlı dinamik bir yapıya dönüştürmektedir.

Bu yapı sayesinde, kullanıcıya sunulan senaryolar yalnızca teorik simülasyonlar olmaktan çıkmakta; gerçek çalışma ortamı içerisinde deneyimlenen ve anlık geri bildirim ile desteklenen öğrenme deneyimlerine dönüşmektedir. Sistem, kullanıcı davranışlarını sürekli izleyerek riskli etkileşimleri anında tespit etmekte ve bu etkileşimleri eğitsel bir fırsata çevirerek mikro-öğrenme modülleri aracılığıyla eğitim sağlamaktadır. Böylece öğrenme süreci zamana yayılan statik bir faaliyet olmaktan çıkarak, olay anına entegre edilmiş bağlamsal bir deneyim haline gelmektedir.

Ayrıca kullanıcı tepkilerinin sisteme geri beslenmesi, modelin zaman içerisinde daha doğru risk tahminleri yapabilmesini ve daha etkili senaryolar üretebilmesini sağlamaktadır. Bu durum, önerilen modelin yalnızca bir eğitim aracı değil, aynı zamanda sürekli öğrenen ve gelişen bir siber güvenlik ekosistemi olarak değerlendirilmesini mümkün kılmaktadır.

4) Örnek Olay İncelemesi: Finans Departmanı Simülasyonu

Geliştirilen bu modelin pratikteki işleyişini somutlaştırmak adına Şekil 3'deki senaryo tasarlanmıştır. Kurumun finans departmanında görev yapan "Kullanıcı A", risk profilleme motoru (1. Aşama) tarafından yapılan veri analizi sonucunda mesai saatleri dışında sık işlem yapması ve önceki testlerde zayıf güvenlik refleksleri göstermesi sebebiyle "Yüksek Risk" profiline atanmıştır.

Eğitim Üretim Motoru (2. Aşama), finans departmanının güncel bağlamını analiz ederek Kullanıcı A'ya kurumun üst düzey mali yöneticisinden gelmiş gibi görünen, "Acil: Çeyrek Sonu Vergi Ceza İhbarnamesi (Banka Dekontları Ektedir)" konulu, yüksek inandırıcılığa sahip bir sosyal mühendislik oltalama e-postası üretir. Bu senaryo, kullanıcının görev alanı, iş yoğunluğu ve karar verme baskısı göz önünde bulundurularak dinamik olarak oluşturulmuştur.

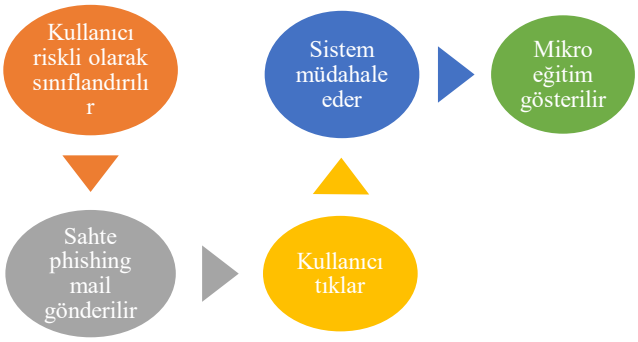
Oluşturulan senaryo, sistem tarafından kullanıcının kurumsal e-posta akışına entegre edilerek iletilir ve kullanıcının etkileşimleri uç nokta izleme ajanı aracılığıyla gerçek zamanlı olarak izlenir. Kullanıcı A, oluşan stres ve aciliyet duygusuyla e-postadaki sahte fatura bağlantısına tıkladığında, bu etkileşim sistem tarafından anlık olarak algılanır.

Tıklama anında uç nokta izleme ajanı devreye girerek potansiyel zararlı işlemi güvenli bir şekilde durdurur ve kullanıcı ekranında kontrollü bir müdahale gerçekleştirir. Sistem, gerçek bir güvenlik ihlalinin oluşmasını engellerken eş zamanlı olarak iCAT mikro-öğrenme modülünü (3. Aşama) başlatır. Kullanıcının ekranında anında beliren yaklaşık 45 saniyelik etkileşimli eğitim içeriği, "aciliyet hissi yaratan mesajların sosyal mühendislikteki rolü" ve "gönderici doğrulama yöntemleri" gibi kritik farkındalık noktalarını açıklamaktadır.

Sunulan bu eğitim içeriği, kullanıcının gerçekleştirdiği hataya doğrudan bağlı olarak oluşturulduğundan, öğrenme süreci bağlamdan kopmadan ve yüksek dikkat düzeyi ile gerçekleşmektedir. Kullanıcı A, teorik bir eğitim yerine, gerçek bir tehdit senaryosu içerisinde doğru güvenlik davranışını deneyimleyerek öğrenmektedir.

Bu süreç sonunda kullanıcı davranışı sistem tarafından kayıt altına alınarak risk profili güncellenir ve elde edilen veriler gelecekte üretilecek senaryoların daha isabetli hale getirilmesi için sisteme geri besleme olarak aktarılır. Böylece model, yalnızca anlık müdahale sağlayan bir yapı olmaktan çıkmakta, sürekli öğrenen ve kullanıcı davranışlarını geliştiren adaptif bir eğitim mekanizmasına dönüşmektedir.

Bu senaryoya ait işlem adımları Şekil 3 üzerinde gösterilmektedir.



Şekil 3. Yapay zekâ destekli siber farkındalık eğitim modelinin senaryo akışı

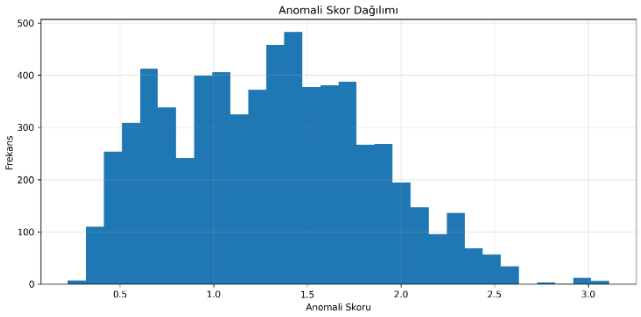
5) Python Tabanlı Simülasyon ve Demo Uygulama

Bu çalışmada önerilen yapay zekâ destekli siber farkındalık modeli, yalnızca teorik bir yaklaşım olarak değil, aynı zamanda gerçek veri üzerinde çalışan bir simülasyon altyapısı ile doğrulanmıştır. Bu kapsamda Python programlama dili kullanılarak geliştirilen sistem, CERT veri seti üzerinde çalıştırılmış ve modelin davranış analizi, risk hesaplama ve senaryo üretim mekanizmaları gerçekçi bir ortamda test edilmiştir.

Simülasyon sürecinde, veri setinde yer alan logon, device, file, email ve http veri kaynakları birlikte işlenerek kullanıcı bazlı bütünleşik bir olay zaman çizelgesi oluşturulmuştur. Bu süreçte veri seti belirli bir zaman aralığına indirgenmiş ve yalnızca seçilen kullanıcılar üzerinden optimize edilerek analiz gerçekleştirilmiştir. Sistem log kayıtlarına göre toplam 6649 adet olay işlenmiş ve her bir kullanıcı için ortalama yaklaşık 1300 olaylık davranış verisi elde edilmiştir. Bu yapı, gerçek kurumsal sistemlerde karşılaşılabilecek yoğunlukta bir veri akışı üzerinde modelin test edilmesini sağlamıştır.

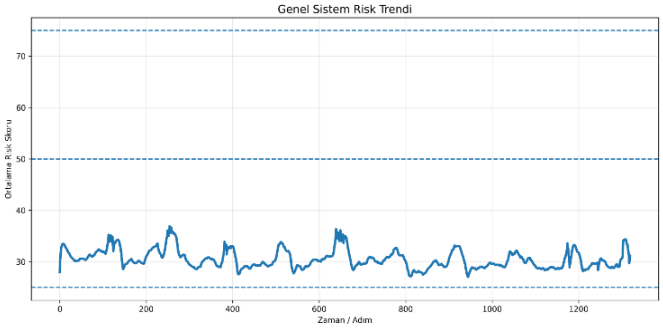
Her bir olay için sistem, kullanıcı davranışını temsil eden çok boyutlu bir gözlem vektörü üretmektedir. Bu vektör; oturum süresi, giriş sayısı, başarısız giriş sayısı, mesai dışı kullanım oranı, veri transfer miktarı, oltalama saldırı tıklama oranı, olay raporlama oranı, cihaz güven skoru ve hassas veri erişim sayısı gibi parametrelerden oluşmaktadır. Bu yaklaşım sayesinde kullanıcı davranışı tek bir metrik üzerinden değil, çok boyutlu bir yapı içerisinde değerlendirilmektedir. Üretilen bu gözlemler, kullanıcının geçmiş davranışlarını temsil eden referans değerleri ile karşılaştırılarak anomali skorları hesaplanmaktadır.

Simülasyon çıktıları incelendiğinde, anomali skorlarının 1.0–2.0 aralığında yoğunlaştığı, bazı durumlarda 2.5 ve üzerine çıktığı görülmektedir.



Şekil 4. Anomali skorlarının dağılımı

Şekil 4 incelendiğinde, anomali skorlarının düşük ve orta seviyelerde yoğunlaştığı, yüksek değerlerin ise sınırlı sayıda uç durumları temsil ettiği görülmektedir. Bu skorların doğrudan risk hesaplama sürecine etki ettiği görülmektedir. Örneğin, belirli bir olayda hesaplanan 2.607 seviyesindeki anomali skoru sonucunda temel risk değeri 36.42, zaman serisi yumuşatma sonrası risk değeri ise 30,95 olarak hesaplanmış ve kullanıcı "orta risk" kategorisinde değerlendirilmiştir. Bu durum, sistemin yalnızca anlık davranış sapmalarını değil, aynı zamanda zamansal trendleri de dikkate aldığını göstermektedir.



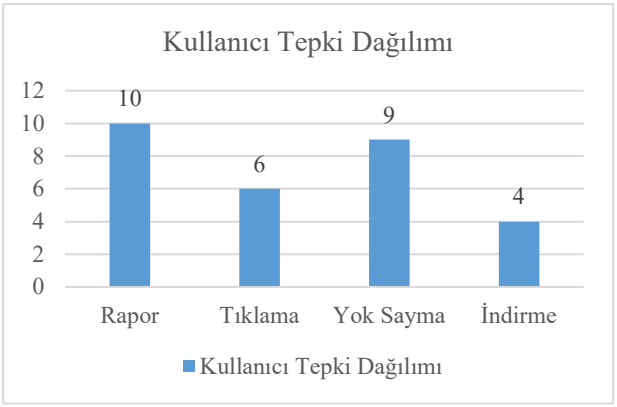
Şekil 5. Sistem risk skorunun zaman içindeki değişimi

Şekil 5, simülasyon süresince hesaplanan sistem risk skorunun zamana bağlı değişimini göstermektedir. Bu şekil, önerilen modelin kullanıcı davranışlarını yalnızca anlık olaylar üzerinden değerlendirmediğini, geçmiş davranışları da dikkate alarak risk seviyesini dinamik biçimde güncellediğini ortaya koymak amacıyla sunulmuştur. Grafikte risk skorunun belirli aralıklar içerisinde dalgalandığı, ancak kritik eşik değerlerini aşmadan kontrol altında kaldığı görülmektedir. Bu durum, sistemin geçici davranış sapmaları ile kalıcı risk eğilimlerini birbirinden ayırt edebildiğini ve risk değerlendirme sürecinde zamansal sürekliliği dikkate aldığını göstermektedir. Ayrıca şekil, önerilen modelin davranış temelli risk izleme mekanizmasının kararlı çalıştığını ve kullanıcı profiline ilişkin değişimleri gerçek zamanlı olarak takip edebildiğini doğrulamaktadır.

Modelin en önemli bileşenlerinden biri olan senaryo tetikleme mekanizması, kullanıcı davranışına duyarlı olarak çalışmaktadır. Senaryo üretimi yalnızca belirli koşullar sağlandığında gerçekleşmektedir. Bu koşullar arasında risk skorunun dinamik eşik değerini aşması, anomali seviyesinin belirli bir minimum değerin üzerinde olması, riskte ani bir artış gözlenmesi ve belirli bir olay aralığında tekrar tetiklenmeyi engelleyen bekleme süresi mekanizması yer

almaktadır. Bu sayede sistem rastgele senaryo üretmek yerine, kullanıcı davranışına bağlı olarak adaptif bir eğitim süreci sunmaktadır.

Simülasyon sonucunda toplam 29 adet siber farkındalık senaryosu tetiklenmiş ve bu senaryolara karşı kullanıcı tepkileri analiz edilmiştir. Kullanıcı davranış dağılımı incelendiğinde, 10 olayda kullanıcıların durumu rapor ettiği, 9 olayda e-postayı yok saydığı, 6 olayda zararlı bağlantıya tıkladığı ve 4 olayda zararlı eki indirdiği görülmektedir. Bu sonuçlar, kullanıcıların farklı risk seviyelerinde ve farklı davranış kalıpları sergilediğini ortaya koymaktadır.



Şekil 6. Kullanıcı tepki türlerinin dağılımı

Şekil 6 incelendiğinde kullanıcıların önemli bir kısmı güvenli davranış (raporlama ve yok sayma) sergilemektedir. Davranışların altında yatan niyet analizi incelendiğinde ise olayların büyük çoğunluğunun kasıtlı değil, "kazara" veya "ihmalkâr" kategorisinde gerçekleştiği belirlenmiştir. Simülasyon sonuçlarına göre 19 olay kazara, 10 olay ise ihmalkâr davranış sonucu meydana gelmiştir. Bu bulgu, literatürde de sıklıkla vurgulanan "iç tehdit" risklerinin büyük ölçüde bilinçsiz kullanıcı davranışlarından kaynaklandığını desteklemektedir.

Senaryo tetiklenen durumlarda sistem, otomatik olarak sosyal mühendislik tabanlı eğitim içerikleri üretmektedir. Bu içerikler; e-posta başlığı, gönderen kimliği, içerik metni, ek dosya türü, saldırı tekniği (MITRE ATT&CK) ve zorluk seviyesi gibi parametreleri içermektedir. Örneğin, simülasyon sırasında üretilen bir senaryoda kullanıcıya "Acil: Çeyrek Sonu Vergi Ceza İhbarnamesi" başlıklı bir oltalama e-posta saldırısı gönderilmiş, bu senaryoda üst yönetim rolü taklit edilmiş ve kullanıcıdan ek dosyayı incelemesi talep edilmiştir. Kullanıcının bu senaryoda eki indirmesi sonucunda risk seviyesi yükselmiş ve sistem tarafından ilgili kullanıcıya mikro eğitim modülü sunulmuştur.

Sistem tarafından üretilen eğitim içerikleri, Bloom taksonomisine uygun olarak yapılandırılmış olup, kullanıcı davranışına göre özelleştirilmektedir. Bu sayede her kullanıcıya aynı eğitim içeriği sunulmak yerine, davranışına özgü farkındalık artırıcı müdahaleler gerçekleştirilmektedir. Bu yaklaşım, klasik statik siber güvenlik eğitimlerinden farklı olarak dinamik ve kişiselleştirilmiş bir öğrenme süreci sağlamaktadır.

Simülasyon genel olarak değerlendirildiğinde, sistemin kullanıcı davranışlarını başarılı bir şekilde analiz ettiği, risk seviyelerini dinamik olarak güncellediği ve yalnızca gerekli durumlarda senaryo üreterek eğitim müdahalesinde bulunduğu gözlemlenmiştir. Ayrıca bazı kullanıcıların yüksek sayıda olay üretmesine rağmen düşük risk seviyesinde kalması, sistemin yalnızca yoğunluk temelli değil, davranış temelli analiz yaptığını göstermektedir.

6) Geliştirilen Modelin Öngörülen Etki Oranı ve Literatür Karşılaştırması

Önerilen bu üç aşamalı yapay zekâ destekli ve mikro-öğrenme tabanlı mimarinin operasyonel başarısı, literatürde gerçekleştirilen kontrollü araştırmaların etki oranları ile tam bir paralellik göstermektedir. Yapılan kapsamlı çalışmalar, siber farkındalık eğitimlerinin kurumlarda oltalama saldırısı kaynaklı zafiyetleri azaltmada önemli bir rol oynadığını göstermektedir. Özellikle kullanıcıların bilgi ve becerilerini geliştirmeyi amaçlayan, etkileşimli ve hedefe yönelik eğitim programlarının, oltalama girişimlerini tanıma, bunlara direnç gösterme ve uygun şekilde raporlama davranışlarını güçlendirdiği belirtilmektedir. Ayrıca düzenli oltalama saldırısı simülasyonları ile desteklenen sürekli farkındalık programlarının, kullanıcıların oltalama saldırılarına karşı daha bilinçli ve dikkatli davranış geliştirmesine katkı sağladığı ortaya konmuştur [46].

Ayrıca, periyodik olarak uygulanan ve simülasyonlarla desteklenen siber farkındalık eğitimlerinin, kullanıcıların sahte e-postalara karşı davranışlarını iyileştirdiği ve riskli etkileşimlerin zaman içerisinde azaldığı gözlemlenmiştir [47]. Güvenlik politikası ihlalleri bağlamında, oyunlaştırma ve mikro-öğrenme temelli farkındalık yaklaşımlarının kullanıcıların güvenlik politikalarına uyumunu artırdığı ve güvenli davranış geliştirmelerini desteklediği belirtilmektedir. Oyunlaştırma tabanlı eğitimlerin, kullanıcı katılımını ve motivasyonunu artırarak güvenlik protokollerine uyumu güçlendirdiği; aynı zamanda bilgi kazanımı ve davranışsal farkındalık üzerinde olumlu etkiler yarattığı deneysel çalışmalarla ortaya konmuştur [48]. Benzer şekilde, mikro-öğrenme ve dürtüleme temelli yaklaşımların kullanıcıların bilgi güvenliği farkındalığını artırdığı, öğrenme performansını iyileştirdiği ve daha güvenli kararlar almaya yönlendirdiği gösterilmiştir [49]. Bununla birlikte, oyunlaştırma temelli siber güvenlik eğitimlerinin kullanıcı motivasyonu, katılımı ve davranışsal değişim üzerinde etkili olabildiği, ancak bu etkinin sürdürülebilirliği ve uzun vadeli davranış değişimi açısından dikkatli tasarım gerektirdiği de literatürde vurgulanmaktadır [50]. Rol bazlı kişiselleştirme ve tehdit simülasyonlarını içeren interaktif eğitim platformlarının, kullanıcıların oltalama e-postalarına karşı daha dikkatli davranmasını sağladığı ve bu tür saldırılara karşı duyarlılığı artırdığı çeşitli çalışmalarla ortaya konmuştur [46].

Literatürde yapılan çalışmalar, AI destekli dinamik senaryoların statik senaryolara kıyasla öğrenci performansını anlamlı ölçüde artırdığını göstermektedir [51]. Bu bulgular, geliştirdiğimiz yapay zekâ destekli otonom senaryo üretim modelinin de benzer şekilde personelin güvenlik farkındalığını artırma potansiyeline sahip olduğunu düşündürmektedir.

IV. Sonuç

Bu çalışma, siber güvenlik ihlallerinde insan faktörünün belirleyici rolünü ele alarak personel kaynaklı siber risklerin teknik zafiyetlerden bağımsız düşünülemeyecek kadar kritik bir güvenlik problemi olduğunu ortaya koymuştur. Literatür incelemesi, kurumlarda yaşanan çok sayıda siber olayın doğrudan ya da dolaylı biçimde çalışan davranışları, farkındalık eksikliği, bilişsel zafiyetler, politika ihlalleri ve sosyal mühendislik saldırılarına açıklık ile ilişkili olduğunu göstermiştir. Bu durum, siber güvenliğin yalnızca teknik önlemlerle sürdürülebilir biçimde sağlanamayacağını; insan davranışlarını merkeze alan bütüncül ve dinamik modellerin gerekli olduğunu açıkça ortaya koymaktadır.

Çalışmada geliştirilen yapay zekâ destekli siber farkındalık eğitim modeli, bu gereksinime cevap vermek amacıyla davranışsal veri toplama, risk profilleme, dinamik senaryo üretimi ve mikro-öğrenme temelli geri bildirim mekanizmalarını tek bir çerçevede bütünleştirmiştir. Önerilen yapı; K-Means ve Mean Shift algoritmalarıyla personelin risk seviyesini belirleyen, Naive Bayes ve doğal dil işleme yaklaşımlarıyla kişiye ve role özgü içerik üreten, iCAT ve Bloom Taksonomisi temelli eğitsel bileşenlerle davranış değişimini destekleyen çok katmanlı bir model sunmaktadır. Bu yönüyle çalışma, klasik ve statik siber farkındalık eğitimlerinden ayrılarak sürekli güncellenebilen, bağlama duyarlı ve uyarlanabilir bir eğitim mimarisi önermektedir.

Özellikle önerilen modelin en güçlü yanı, personeli yalnızca bilgi alan pasif bir kullanıcı olarak değil; davranışsal izleri analiz edilen, risk düzeyi ölçülen ve eğitim içeriği buna göre şekillendirilen aktif bir güvenlik bileşeni olarak ele almasıdır. Bu yaklaşım, siber farkındalık eğitimlerini genel geçer içerik sunumundan çıkarıp, gerçek iş akışıyla ilişkili, yüksek gerçeklik düzeyine sahip ve anlık geri bildirim üreten bir yapıya dönüştürmektedir. Böylece çalışanların yalnızca teorik bilgi edinmesi değil, aynı zamanda tehdit anında doğru refleks geliştirmesi hedeflenmektedir. Finans departmanı örneği üzerinden kurgulanan olay incelemesi de, modelin kurumsal bağlamda nasıl uygulanabileceğini ve kişiselleştirilmiş senaryolarla davranışsal öğrenmeyi nasıl güçlendirebileceğini göstermektedir.

Araştırmanın bulguları, yapay zekâ destekli, simülasyon temelli ve mikro-öğrenme odaklı eğitim modellerinin; personelin oltalama saldırılarına karşı duyarlılığını artırma, güvenlik politikalarına uyumu güçlendirme ve insan kaynaklı siber riskleri azaltma bakımından önemli bir potansiyele sahip olduğunu göstermektedir. Bu nedenle kurumların siber güvenlik stratejilerini yalnızca ağ güvenliği, şifreleme veya saldırı tespit sistemleri gibi teknik katmanlarla sınırlamaması; insan davranışlarını ölçen, analiz eden ve eğiten sistemlerle desteklemesi gerekmektedir. Kurumsal dirençliliğin artırılması açısından en etkili yaklaşım, teknik savunma altyapısı ile insan odaklı farkındalık mekanizmalarının eş zamanlı ve veri temelli biçimde işletilmesidir.

Bununla birlikte, önerilen model kavramsal ve kuramsal bir çerçeve sunmakta olup, farklı sektörlerde gerçek kullanıcı verileriyle test edilmesi gelecek çalışmalar açısından önem taşımaktadır. Özellikle sağlık, finans, kamu ve eğitim gibi kritik sektörlerde yapılacak uygulamalı çalışmalar; modelin doğruluğunu, ölçeklenebilirliğini ve uzun vadeli davranış değişimi üzerindeki etkisini somut biçimde ortaya koyacaktır. Ayrıca gelecekte yapılacak araştırmalarda etik veri toplama, çalışan mahremiyeti, açıklanabilir yapay zekâ mekanizmaları ve kurumsal kabul düzeyi gibi konuların da modele entegre edilmesi gerekmektedir.

Sonuç olarak bu çalışma, personel kaynaklı siber risklerin azaltılmasına yönelik yapay zekâ destekli yeni bir siber farkındalık modeli ortaya koymaktadır. Model, çalışan davranışlarını analiz ederek kişiye özel eğitim süreçleri üretmektedir. Çalışma, teknik güvenlik yaklaşımlarını insan davranış analitiği ile bir araya getiren bütüncül bir yapı sunmaktadır. Geleceğin siber güvenlik anlayışında başarılı kurumlar, yalnızca sistemlerini değil çalışanlarının karar verme süreçlerini de koruyabilen kurumlar olacaktır. Adaptif ve sürekli öğrenebilen siber farkındalık sistemlerinin kurumsal siber savunmanın temel unsurlarından biri haline geleceği öngörülmektedir.

V. Ek Kaynak ve Materyal

Bu araştırma kapsamında geliştirilen Python kodları ve ilgili ek materyaller, çalışmanın şeffaflığı ve tekrar üretilebilirliği ilkesi gereğince çevrimiçi bir platformda erişime açılacaktır. İlgili okuyucular ve araştırmacılar, deney düzeneklerine, kaynak kodlara ve metodolojinin birebir tekrarı için gerekli tüm tamamlayıcı içeriklere, kaynaklar bölümünde atıfta bulunulan [52] numaralı bağlantı aracılığıyla ulaşabilirler. Söz konusu dijital materyallerin paylaşılması, elde edilen bulguların doğrulanabilirliğini ve araştırmanın yeniden üretilebilirliğini desteklemeyi hedeflemektedir.

References

[1] "Security Awareness Training for End Users on a Limited Budget."

[2] S. Colabianchi, F. Costantino, F. Nonino, and G. Palombi, "Transforming threats into opportunities: The role of human factors in enhancing cybersecurity," *Journal of Innovation and Knowledge*, vol. 10, no. 3, May 2025, doi: 10.1016/j.jik.2025.100695.

[3] Sudha Rani Pujari and Mahmood Afzal Hussain, "Human Factor in Cybersecurity: Behavioral Insights into Phishing and Social Engineering Attacks," *Nanotechnol. Percept.*, pp. 630–642, Dec. 2024, doi: 10.62441/NANO-NTP.VI.3556.

[4] K. Khando, S. Gao, S. M. Islam, and A. Salman, "Enhancing employees information security awareness in private and public organisations: A systematic literature review," *Comput. Secur.*, vol. 106, Jul. 2021, doi: 10.1016/j.cose.2021.102267.

[5] T. Longtchi, R. M. Rodriguez, L. Al-Shawaf, A. Atyabi, and S. Xu, "Internet-based Social Engineering Attacks, Defenses and Psychology: A Survey," Aug. 2022, [Online]. Available: http://arxiv.org/abs/2203.08302

[6] C. GÜNDÜZALP, "Üniversite Çalışanlarının Dijital Veri ve Kişisel Siber Güvenlik Farkındalıkları (Bilgi İşlem Daire Başkanlıkları Örneği)," *Journal of*

Computer and Education Research, vol. 9, no. 18, pp. 598–625, Dec. 2021, doi: 10.18009/jcer.907022.

[7] A. A. Darem, A. A. Alhashmi, T. M. Alkhaldi, A. M. Alashjaee, S. M. Alanazi, and S. A. Ebad, "Cyber Threats Classifications and Countermeasures in Banking and Financial Sector," *IEEE Access*, vol. 11, pp. 125138–125158, 2023, doi: 10.1109/ACCESS.2023.3327016.

[8] S. M. Ali, A. Razzaque, M. Yousaf, and R. U. Shan, "An Automated Compliance Framework for Critical Infrastructure Security Through Artificial Intelligence," *IEEE Access*, vol. 13, pp. 4436–4459, 2025, doi: 10.1109/ACCESS.2024.3524496.

[9] M. Auz-Torres and S. G. Yoo, "Key Factors for Reducing Cybersecurity Risks in Healthcare: A Systematic Literature Review," *IEEE Access*, vol. 14, pp. 33002–33033, 2026, doi: 10.1109/ACCESS.2026.3667073.

[10] A. Ali, C. Overboe, T. Mohottalalage, and A. B. Mailewa, "The Human Factor in Cybersecurity: Understanding Psychology, Training Efficacy, and Error Reduction Strategies."

[11] C. Luca, J. Miller, and M. Grant, "Statistical Risk Assessment of Human Factor Vulnerabilities in Medical Cybersecurity."

[12] T. Rathod, N. K. Jadav, S. Tanwar, A. Alabdulatif, D. Garg, and A. Singh, "A comprehensive survey on social engineering attacks, countermeasures, case study, and research challenges," *Inf. Process. Manag.*, vol. 62, no. 1, Jan. 2025, doi: 10.1016/j.ipm.2024.103928.

[13] C. A. Bilen, "Are critical success factors for cybersecurity just technical issues Cybersecurity management and the human factor," *Pressacademia*, Jun. 2023, doi: 10.17261/pressacademia.2023.1735.

[14] Diana Ussher-Eke, "From awareness to action: Designing effective cybersecurity training programs," *International Journal of Science and Research Archive*, vol. 16, no. 2, pp. 494–504, Aug. 2025, doi: 10.30574/ijsra.2025.16.2.2348.

[15] "Assessing the Efficacy of Security Awareness Training in Mitigating Phishing Attacks: A Review," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 14, no. 3, pp. 177–184, Jun. 2025, doi: 10.30534/ijatcse/2025/081432025.

[16] Victoria Bukky Ayoola, Ugoaghalam Uche James, Idoko Peter Idoko, Onuh Matthew Ijiga, and Toyosi Motilola Olola, "Effectiveness of social engineering awareness training in mitigating spear phishing risks in financial institutions from a cybersecurity perspective," *Global Journal of Engineering and Technology Advances*, vol. 20, no. 3, pp. 094–117, Sep. 2024, doi: 10.30574/gjeta.2024.20.3.0164.

[17] M. T. Jafar, W. Ren, L. X. Yang, and G. Li, "Attack Graph-Epidemic Hybrid Modeling for Mitigating Cyber Threats Propagation," *IEEE Trans. Comput. Soc. Syst.*, 2026, doi: 10.1109/TCSS.2025.3635013.

[18] İ. TUĞAL, C. ALMAZ, and M. SEVİ, "Üniversitelerdeki Siber Güvenlik Sorunları ve Farkındalık Eğitimleri," *Bilişim Teknolojileri Dergisi*, vol. 14, no. 3, pp. 229–238, Jul. 2021, doi: 10.17671/gazibtd.754458.

[19] A. EFE, "Yapay Zeka Odaklı Siber Risk ve Güvenlik Yönetimi," *Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi*, vol. 5, no. 2, pp. 144–165, Dec. 2021, doi: 10.33461/uybisbbd.972206.

[20] J. Zywiołek, M. Matulewski, and J. Frąś, "AI as a shield against cyberattacks - Employee awareness in the EU," in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 5510–5519. doi: 10.1016/j.procs.2025.10.019.

[21] R. Al-Kuwari, "Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000. Gamified Cybersecurity Awareness Training: A Framework and Evaluation", doi: 10.1109/ACCESS.2017.DOI.

[22] J. Prümmer, T. van Steen, and B. van den Berg, "Assessing the effect of cybersecurity training on End-users: A Meta-analysis," *Comput. Secur.*, vol. 150, Mar. 2025, doi: 10.1016/j.cose.2024.104206.

[23] E. A. Kenlie, C. N. Hendris, N. Nadia, and F. Adeta, "The Impact of Human Error in Cybersecurity Breaches: Understanding Behavioral Patterns and Mitigation Strategies," in *Proceedings - 6th International Conference on Informatics, Multimedia, Cyber and Information System, ICIMCIS 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 1045–1050. doi: 10.1109/ICIMCIS63449.2024.10956680.

[24] M. Bada, A. M. Sasse, and J. R. C. Nurse, "Cyber Security Awareness Campaigns: Why do they fail to change behaviour?"

[25] H. A. Dewi, C. Candiwan, and P. K. Sari, "Artificial Intelligence in Security Education, Training and Awareness: A Bibliometric Analysis," in *2024 International Conference on Intelligent Cybernetics Technology and Applications, ICICyTA 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 914–919. doi: 10.1109/ICICYTA64807.2024.10912940.

[26] "Insider Threat Test Dataset." Accessed: Apr. 22, 2026. [Online]. Available: https://www.sei.cmu.edu/library/insider-threat-test-dataset/

[27] M. Mohapatra, A. Phukan, and V. K. Madisetti, "Generative Adversarial Network Based Approach towards Synthetically Generating Insider Threat Scenarios," *Journal of Software Engineering and Applications*, vol. 16, no. 11, pp. 586–604, 2023, doi: 10.4236/jsea.2023.1611030.

[28] G. ASLANTAŞ, M. GENÇGÜL, M. RUMELLİ, M. ÖZSARAÇ, and G. BAKIRLI, "Customer Segmentation Using K-Means Clustering Algorithm and RFM Model," *Deu Muhendislik Fakultesi Fen ve Muhendislik*, vol. 25, no. 74, pp. 491–503, May 2023, doi: 10.21205/deufmd.2023257418.

[29] C. Cariou, S. Le Moan, and K. Chehdi, "A Novel Mean-Shift Algorithm for Data Clustering," *IEEE Access*, vol. 10, pp. 14575–14585, 2022, doi: 10.1109/ACCESS.2022.3147951.

[30] J. Webb, F. Abri, and S. Akther, "Synthetic Social Engineering Scenario Generation Using LLMs for Awareness-Based Attack Resilience," *IEEE Access*,

vol. 13, pp. 174831–174856, 2025, doi: 10.1109/ACCESS.2025.3614550.

[31] N. Konyrbaev *et al.*, "EVALUATION AND OPTIMIZATION OF THE NAIVE BAYES ALGORITHM FOR INTRUSION DETECTION SYSTEMS USING THE USB-IDS-1 DATASET," *Eastern-European Journal of Enterprise Technologies*, vol. 6, no. 2(132), pp. 74–82, 2024, doi: 10.15587/1729-4061.2024.317471.

[32] D. Üniversitesi and M. Gül TAŞ, "Semantic Detection of Turkish Phishing Emails: A Natural Language Processing and Deep Learning-Based Approach," 2025. [Online]. Available: https://dergipark.org.tr/tr/pub/cdej

[33] K. Sarpong Adu-Manu, R. Kwasi Ahiable, J. Kwame Appati, and E. Essel Mensah, "Phishing Attacks in Social Engineering: A Review," *Journal of Cyber Security*, vol. 4, no. 4, pp. 239–267, 2022, doi: 10.32604/jcs.2023.041095.

[34] S. Roy, E. Panaousis, C. Noakes, A. Laszka, S. Panda, and G. Loukas, "SoK: The MITRE ATT&CK Framework in Research and Practice," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.07411

[35] H. Taherdoost, "Towards an Innovative Model for Cybersecurity Awareness Training," *Information (Switzerland)*, vol. 15, no. 9, Sep. 2024, doi: 10.3390/info15090512.

[36] F. Ş. Fayda Kınık and A. Kirişçi Sarıkaya, "Online learning at the post-graduate level: Interpretations through Bloom's revised taxonomy," *Turkish Journal of Education*, vol. 14, no. 1, pp. 67–92, Jan. 2025, doi: 10.19128/turje.1454562.

[37] F. L. Greitzer and R. E. Hohimer, "Modeling Human Behavior to Anticipate Insider Attacks," *Journal of Strategic Security*, vol. 4, no. 2, pp. 25–48, Jun. 2011, doi: 10.5038/1944-0472.4.2.2.

[38] L. Koli, S. Kalra, R. Thakur, A. Saifi, and K. Singh, "AI-Driven IRM: Transforming insider risk management with adaptive scoring and LLM-based threat detection," May 2025, [Online]. Available: http://arxiv.org/abs/2505.03796

[39] L. Schöni, V. Carles, M. Strohmeier, P. Mayer, and V. Zimmermann, "You Know What? - Evaluation of a Personalised Phishing Training Based on Users' Phishing Knowledge and Detection Skills," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Nov. 2024, pp. 1–14. doi: 10.1145/3688459.3688460.

[40] F. Greco, G. Desolda, C. Tucci, A. Esposito, A. Curci, and A. Piccinno, "Improving Phishing Resilience with AI-Generated Training: Evidence on Prompting, Personalization, and Duration," Dec. 2025, [Online]. Available: http://arxiv.org/abs/2512.01893

[41] P. Sirawongphatsara, P. Pornpongtechavanich, N. Phanthuna, and T. Daengsi, "Comparative Simulation of Phishing Attacks on a Critical Information Infrastructure Organization: An Empirical Study," *Bulletin of Electrical Engineering and Informatics*, vol. 99, no. 1, pp. 1–1, 2099, doi: 10.11591/eei.v99i1.paperID.

[42] D. Lain, T. Jost, S. Matetic, K. Kostiainen, and S. Capkun, "Content, Nudges and Incentives: A Study on the Effectiveness and Perception of Embedded Phishing Training," in *CCS 2024 - Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security*, Association for Computing Machinery, Inc, Dec. 2024, pp. 4182–4196. doi: 10.1145/3658644.3690348.

[43] F. M. Cau, G. Desolda, F. Greco, L. D. Spano, and L. Viganò, "Can Large Language Models Automate Phishing Warning Explanations? A Controlled Experiment on Effectiveness and User Perception," Dec. 2025, [Online]. Available: http://arxiv.org/abs/2507.07916

[44] N. D. Tantaroudas, I. Karachalios, and A. J. McCracken, "SentinelSphere: Integrating AI-Powered Real-Time Threat Detection with Cybersecurity Awareness Training," Apr. 2026, [Online]. Available: http://arxiv.org/abs/2604.06900

[45] X. Chen, M. Sacre, G. Lenzini, S. Greiff, V. Distler, and A. Sergeeva, "The Effects of Group Discussion and Role-playing Training on Self-efficacy, Support-seeking, and Reporting Phishing Emails: Evidence from a Mixed-design Experiment," in *Conference on Human Factors in Computing Systems - Proceedings*, Association for Computing Machinery, May 2024. doi: 10.1145/3613904.3641943.

[46] F. Iqbal *et al.*, "Efficacy of Cybersecurity Awareness Training in Reducing Phishing Vulnerabilities in Organizations," 2024. [Online]. Available: https://polarpublications.com/index.php/JACSTIC/article/view/2

[47] N. Marshall, D. Sturman, and J. C. Auton, "Exploring the evidence for email phishing training: A scoping review," *Comput. Secur.*, vol. 139, Apr. 2024, doi: 10.1016/j.cose.2023.103695.

[48] M. Research, C. Efoseh, I. Fachano, and V. Ayala-Rivera, "Gamification in Cybersecurity: Improving Employee Compliance through Game-Based Learning and Incentives."

[49] K. Hiraoka, S. Matsuo, Y. Arakawa, and Y. Nakamura, "WiLearn: Design and Implementation of a Microlearning System that Utilizes a Captive Portal of Wi-Fi," *2023 14th International Conference on Mobile Computing and Ubiquitous Network, ICMU 2023*, 2023, doi: 10.23919/ICMU58504.2023.10412210.

[50] A. K. Gwenhure and S. G. Nam, "GAMIFIED CYBERSECURITY INITIATIVES: THE TREND, LIMITS, AND LESSONS," *Journal of Information Technology Education: Research*, vol. 24, pp. 1–34, 2025, doi: 10.28945/5601.

[51] Dr. S. Thapaliya and Mr. D. Adhikari, "The Future of AI in Cybersecurity Education: Training the Next Generation – How AI can be Used to Teach and Train Students in Cybersecurity Fields," *International Journal of Multidisciplinary and Innovative Research*, vol. 02, no. 04, Apr. 2025, doi: 10.58806/ijmir.2025.v2i4n01.

[52] "sinanozcelikk/cyber-risk-training-code · GitHub." Accessed: Apr. 22, 2026. [Online]. Available:

https://github.com/sinanozcelikk/cyber-risk-training-code

IAGOA: An Improved Adaptive Grasshopper Optimization Algorithm with Superior Stability for Mobile Robot Path Planning

Tohid Yousefi
Faculty of Computer and Information Technologies
Cappadocia University
Nevsehir/Turkey
tohid.yousefi@kapadokya.edu.tr

***Abstract*— Mobile robot path planning in complex environments remains challenging due to non-convex constraints, collision risks, and the need for smooth trajectories. While the Grasshopper Optimization Algorithm (GOA) has shown promise, existing variants suffer from premature convergence, low stability, and inconsistent performance. This paper proposes IAGOA (Improved Adaptive Grasshopper Optimization Algorithm), integrating five key mechanisms: (i) chaotic Tent-map initialization with Opposition-Based Learning, (ii) adaptive Lévy Flight with dynamic step size adjustment, (iii) Gbest-Guided Differential Mutation, (iv) Greedy Position Inheritance, and (v) Dynamic Population Restart. Extensive experiments benchmark IAGOA against GOA, AGOA, PSO, GWO, and GA using Wilcoxon and Friedman tests. Results demonstrate that IAGOA significantly outperforms all competitors, achieving the lowest mean cost with exceptional stability—an order-of-magnitude improvement in reliability. Wilcoxon tests confirm statistical superiority at α=0.05 against every comparator, including GWO, while an extremely narrow best-worst range proves exceptional run-to-run consistency. The Friedman test secures IAGOA as the top-performing algorithm.**

Keywords*— *Grasshopper Optimization Algorithm, Mobile Robot Path Planning, Metaheuristic Optimization, Adaptive Lévy Flight, Opposition-Based Learning

I. INTRODUCTION

Autonomous mobile robots are indispensable in modern industrial and service sectors, from warehouse logistics to search-and-rescue operations [1]. A central challenge in these domains is path planning: generating a collision-free, optimal trajectory from a start to a goal configuration within a constrained environment [2]. This problem is typically NP-hard, and its complexity grows exponentially with the dimensionality of the workspace and the number of obstacles [3]. Classical approaches like the A* algorithm and Dijkstra's method guarantee completeness but become computationally prohibitive in high-dimensional or dynamic spaces. Sampling-based planners such as Rapidly-exploring Random Trees (RRT) offer probabilistic completeness but produce jerky, suboptimal paths requiring post-processing [4].

To address the shortcomings of deterministic and sampling-based methods, a significant body of research has turned to metaheuristic optimization algorithms. These population-based methods, inspired by biological and physical phenomena, do not require gradient information and are adept at escaping local optima in non-convex search landscapes [5]. Among these, Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) have been extensively applied but frequently suffer from premature convergence and high computational cost, respectively [6]. The Grey Wolf Optimizer (GWO), inspired by the strict social hierarchy of grey wolves, offers a better balance between exploration and exploitation but can still stagnate in complex, multi-obstacle environments, as evidenced by the high standard deviation in our experiments. The Grasshopper Optimization Algorithm (GOA) [7], which mathematically models the social interaction of grasshopper swarms, has emerged as a powerful alternative. Its dual-phase mechanism—a repulsion zone for exploration and an attraction zone for exploitation—makes it particularly apt for path planning. However, the canonical GOA exhibits a critical limitation: the linear decay of its comfort zone parameter prematurely shifts the swarm from exploration to exploitation, leading to entrapment in local optima and a high variance in solution quality across runs [8].

To overcome these fundamental limitations, several adaptive variants have been proposed. The Adaptive GOA (AGOA) attempts to dynamically tune the comfort zone parameter, yet our empirical analysis shows this provides marginal improvement, with stability and mean cost remaining unsatisfactory (see Table I). This gap motivates the core contribution of this paper: the **I**mproved **A**daptive **G**rasshopper **O**ptimization **A**lgorithm (IAGOA). We propose a synergistic suite of five novel, integrated mechanisms that fundamentally enhance the algorithm's exploratory capacity, exploitative precision, and robustness: (i) a chaotic Tent-map initialization fused with Opposition-Based Learning to maximize initial population diversity, (ii) an adaptive Lévy Flight mechanism with dynamic step size adjustment to effectively escape local minima without compromising late-stage convergence, (iii) a Gbest-Guided Differential Mutation to accelerate exploitation of the most promising search regions, and (iv) a Greedy Position Inheritance strategy to exploit local neighborhood knowledge, and (v) a Dynamic Population Restart mechanism to maintain swarm vitality and prevent stagnation.

We conduct a rigorous and comprehensive benchmark of IAGOA against the canonical GOA, its direct competitor AGOA, and the widely adopted PSO, GWO, and GA. The extensive evaluation uses non-parametric statistical tests, including the Wilcoxon signed-rank test for pairwise comparisons and the Friedman test for global rank ordering, at a 0.05 significance level. The results unequivocally demonstrate the superiority of IAGOA. It achieves the lowest mean cost with the smallest standard deviation, representing an order-of-magnitude improvement in reliability. Critically, IAGOA is statistically proven to outperform every

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

comparator, including the previously best-performing GWO, with its run-to-run consistency validated by an extremely narrow best-worst range on the complex path planning problem modeled in this study.

The remainder of this paper is structured as follows. Section II formulates the path planning problem and details the objective function. Section III presents the canonical GOA and elaborates on the five key innovations of the proposed IAGOA. Section IV discusses the experimental setup, the benchmark results, and the statistical analyses. Finally, Section V concludes the paper and suggests promising avenues for future research.

II. PROBLEM FORMULATION AND MODELING

A. Path Planning as an Optimization Problem

The objective of mobile robot path planning is to determine a continuous, collision-free path $P = \{S, p_1, p_2, \dots, p_n, T\}$ from a designated start point $S$to a target point T within a bounded 2D workspace $W \subset \mathbb{R}^2$. The workspace is populated with a set of $m$static obstacles $O = \{O_1, O_2, \dots, O_m\}$, each defined by its centroid coordinates (o_x^k, o_y^k) and radius r_k. We model the path using $n$intermediate waypoints, yielding a $2n$-dimensional decision vector $\mathrm{X} = [x_1, \dots, x_n, y_1, \dots, y_n]^T$, where each pair (x_i, y_i) represents a waypoint constrained within $[x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]$. A cubic spline interpolation through the sequence $\{S, (x_1, y_1), \dots, (x_n, y_n), T\}$ generates a C^2 -continuous trajectory comprising $K = 200$ uniformly sampled points, ensuring smooth and kinematically executable paths [9]. The path planning problem is formulated as a constrained single-objective minimization over the feasible space $\mathcal{F} \subset \mathbb{R}^{2n}$:

$$\min_{\mathrm{X} \in \mathcal{F}} \; F(\mathrm{X}) = L(\mathrm{X}) + \lambda_1 S(\mathrm{X}) + \lambda_2 V(\mathrm{X})$$

where $F(\mathrm{X})$ is the aggregate cost function, $L(\mathrm{X})$ is the path length, $S(\mathrm{X})$ is the smoothness penalty, and $V(\mathrm{X})$ is the collision violation term. The weighting coefficients $\lambda_1 = 0.1$and $\lambda_2 = 10^6$ are calibrated to balance path optimality with safety, imposing a dominant penalty on any obstacle infringement [2, 10].

B. Objective Function Components

1) Path Length $L(\mathrm{X})$: The total Euclidean distance traversed by the robot is computed as the sum of segment lengths between consecutive interpolated trajectory points:

$$L(\mathrm{X}) = \sum_{i=1}^{K-1} \sqrt{\left(x_{i+1}^p - x_i^p\right)^2 + \left(y_{i+1}^p - y_i^p\right)^2}$$

where (x_i^p, y_i^p) denotes the i-th point on the cubic spline trajectory. Minimizing $L(\mathrm{X})$directly reduces travel time and energy consumption [2].

2) Smoothness Penalty $S(\mathrm{X})$: Sharp turns produce kinematically infeasible trajectories for nonholonomic mobile robots and cause excessive wear on actuators. The smoothness cost penalizes cumulative changes in heading angle along the path:

$$S(\mathrm{X}) = \sum_{i=1}^{K-2} \left| \arctan 2(\Delta y_{i+1}, \Delta x_{i+1}) - \arctan 2(\Delta y_i, \Delta x_i) \right|$$

where $\Delta x_i = x_{i+1}^p - x_i^p$ and $\Delta y_i = y_{i+1}^p - y_i^p$ are the coordinate differences between consecutive spline points. Lower values of $S(\mathrm{X})$ correspond to smoother, more executable trajectories [10].

3) Collision Violation $V(\mathrm{X})$: Safety is the primal constraint in autonomous navigation. The collision penalty quantifies the degree of obstacle infringement using a normalized distance function averaged over all trajectory points and all obstacles:

$$V(\mathrm{X}) = \frac{1}{K \cdot m} \sum_{i=1}^{K} \sum_{k=1}^{m} \max\left(0, 1 - \frac{d_{i,k}}{r_k}\right)$$

where $d_{i,k} = \sqrt{\left(x_i^p - o_x^k\right)^2 + \left(y_i^p - o_y^k\right)^2}$is the Euclidean distance from the i-th path point to the center of the k-th obstacle, and r_k is the obstacle radius. The term is zero when the path maintains at least distance r_k from all obstacle centers and increases linearly as the path penetrates an obstacle, scaled by $\lambda_2 = 10^6$ to effectively forbid any collision [1].

C. Search Space Complexity

The cost landscape defined by $F(\mathrm{X})$ is highly non-convex and riddled with local minima due to the superimposed collision penalty fields surrounding each obstacle. The six-obstacle workspace creates narrow corridors between start point $S = (0,0)$ and goal point $T = (10,10)$, forcing algorithms to navigate through constrained passages while optimizing path length and smoothness. This multi-modal landscape serves as an ideal testbed for evaluating the exploration-exploitation balance of metaheuristic optimizers.

III. THE PROPOSED IAGOA METHODOLOGY

The canonical Grasshopper Optimization Algorithm (GOA) updates the position of the i-th grasshopper in the d-th dimension as [7]:

$$X_i^d(t+1) = c\left(\sum_{j=1, j\neq i}^{N} c\frac{ub_d - lb_d}{2} s(\left| x_j^d - x_i^d \right|)\frac{x_j - x_i}{d_{ij}}\right) + \hat{T}_d$$

where $s(r) = Fe^{-r/l} - e^{-r}$defines social forces, and the comfort zone parameter $c$decreases linearly from $c_{\max} = 1.0$ to $c_{\min} = 0.00004$. This linear decay prematurely shifts the swarm from exploration to exploitation, causing stagnation in local minima and high run-to-run variance [8]. IAGOA overcomes these limitations through five synergistic mechanisms.

Mechanism 1 (Chaotic Tent-Map Initialization with OBL): Initial population quality critically affects convergence. IAGOA first generates a chaotic Tent map sequence $x_{k+1} = 2x_k$ if $x_k < 0.5$, else $2(1 - x_k)$, providing superior ergodicity over random initialization [3]. Opposition-Based Learning then computes $\bar{\mathrm{X}} = \mathrm{lb} + \mathrm{ub} - \mathrm{X}$ for each individual. The N fittest from the combined $2N$ pool form the initial swarm, guaranteeing maximum diversity.

Mechanism 2 (Adaptive Lévy Flight with Dynamic Step Size): To escape local optima, the social interaction is augmented with Lévy flight perturbation. The step size adapts dynamically based on the global best improvement rate $\delta(\mathcal{H}) = \left| f_{t-4} - f_t \right| / (\left| f_{t-4} \right| + \epsilon)$, computed over a five-iteration window: $\alpha(t) = \alpha_0 / (1 + 10 \cdot \delta(\mathcal{H}))$. Stagnation

increases $\alpha(t)$, enabling long jumps via the heavy-tailed Lévy distribution ($\beta = 1.5$); rapid progress shrinks it for refinement. The perturbation $\alpha(t) \cdot \text{Lévy}(\beta) \odot (\mathrm{X}_i - \widehat{\mathrm{T}})$ is retained only if fitness improves.

Mechanism 3 (Gbest-Guided Differential Mutation): A mutation vector $\mathrm{V}_i = \widehat{\mathrm{T}} + F(t) \cdot (\mathrm{X}_{r1} - \mathrm{X}_{r2})$ is generated, where $F(t) = 0.3 + 0.4(1 - t/T)$ decays adaptively. Binomial crossover produces a trial vector that replaces X_i if fitter, systematically pulling the swarm toward promising regions.

Mechanism 4 (Greedy Position Inheritance): When the global best stagnates for five iterations, each individual moves toward its fittest neighbor within a k-nearest neighborhood ($k = \lfloor N/5 \rfloor$): $\mathrm{X}_i^{new} = \mathrm{X}_i + \eta(t) \cdot (\mathrm{X}_{nb}^{*} - \mathrm{X}_i)$, with $\eta(t) = 0.08(1 - t/T)$. This local information sharing complements global guidance.

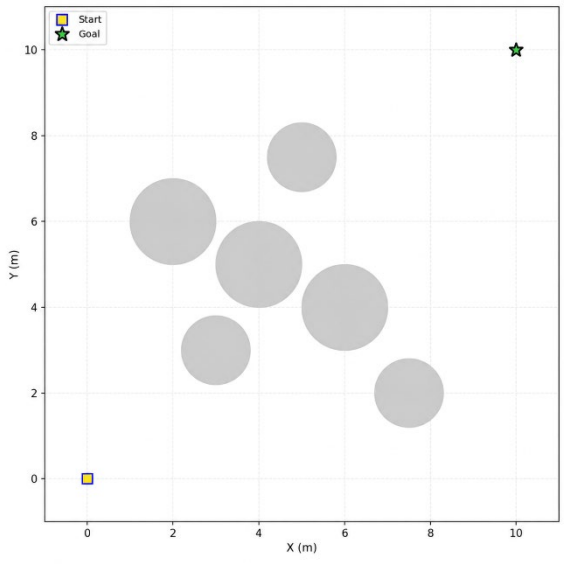
Mechanism 5 (Dynamic Population Restart): Upon 15 iterations of stagnation, the worst 50% of the population is regenerated via chaotic Tent-map initialization with OBL, while the elite half is preserved. This injects exploratory energy while retaining acquired knowledge, breaking deep stagnation traps.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Setup

All algorithms were evaluated on a $10 \times 10\text{m}^2$ workspace with six circular obstacles forming narrow corridors between $S = (0,0)$ and $T = (10,10)$, *as illustrated in Fig.* 1. The path is modeled with $n = 5$ waypoints, yielding a 10-dimensional optimization problem. Each algorithm executed 30 independent runs with population size $N = 30$ and maximum iterations $T = 100$. Experiments were conducted in Python 3.x, and statistical analyses were performed using SciPy at significance level $\alpha = 0.05$. IAGOA is benchmarked against GOA [7], AGOA , PSO [11], GWO [12], and GA [13].

Fig. 1. Robot path planning environment with obstacles, start, and goal positions.



B. Solution Quality and Stability

Table I summarizes the performance metrics across all algorithms. IAGOA achieves the lowest mean cost (16.27) with exceptional stability ($\sigma = 0.28$), representing a 62.3× reduction in standard deviation compared to PSO ($\sigma = 64.27$) and a 23.3× reduction compared to GWO ($\sigma = 6.52$). The best-worst range of IAGOA is merely 1.43 units ($16.49 - 15.06$), demonstrating near-deterministic run-to-run consistency unmatched by any competitor. While AGOA achieves a marginally better best cost (15.09) than IAGOA (15.06), its worst cost (23.99) and standard deviation (1.83) reveal significant instability, with a best-worst span of 8.90 units—6.2× wider than IAGOA. PSO and GA exhibit catastrophic worst-case performance (377.66 and 5770.20, respectively), confirming their unsuitability for reliable autonomous navigation where a single collision is unacceptable.

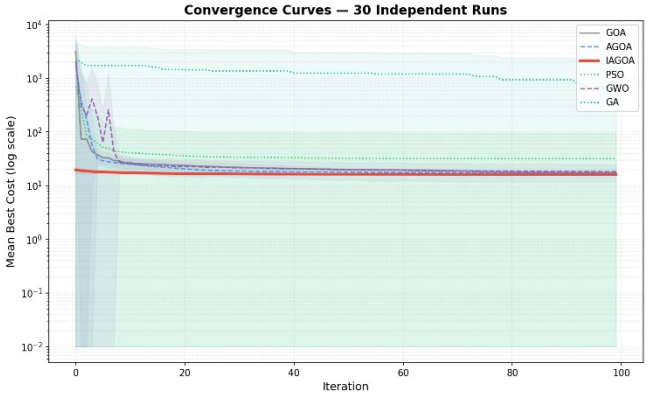
Table I. Comparative Performance Across 30 Independent Runs

Algorithm	Mean	Std	Best	Worst
GOA	17.35	0.99	15.60	20.66
AGOA	17.59	1.83	15.09	23.99
IAGOA	**16.27**	**0.28**	15.06	**16.49**
PSO	31.92	64.27	16.38	377.66
GWO	18.30	6.52	15.50	49.77
GA	670.60	1284.84	16.69	5770.20

C. Convergence Analysis

Fig. 2 illustrates the mean convergence curves over 30 runs. IAGOA exhibits rapid initial descent, converging to near-optimal solutions within the first 20 iterations—significantly faster than all competitors. The logarithmic scale exposes the divergent behavior of PSO and GA, whose curves oscillate wildly due to repeated collision with obstacles. GWO converges steadily but plateaus around iteration 60 at a suboptimal cost level, indicating premature stagnation. Critically, the shaded standard deviation region of IAGOA remains the narrowest throughout the entire optimization process, confirming the algorithm's robust performance across all runs.

Fig. 2. Mean convergence curves with standard deviation shading (30 runs)

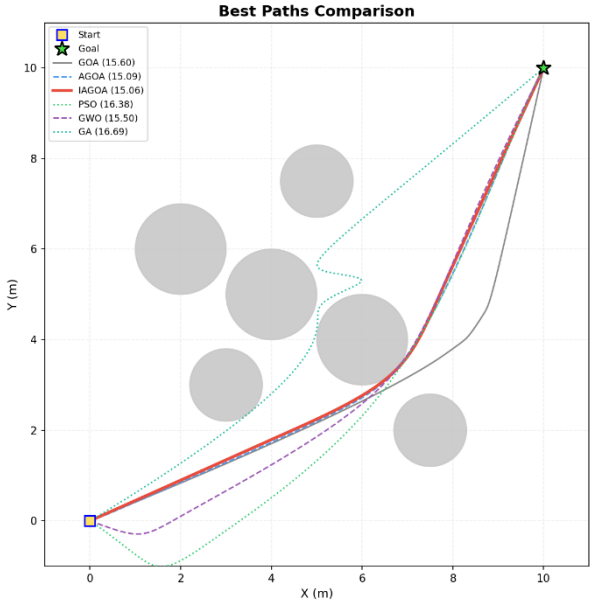


D. Path Quality Visualization

Fig. 3 displays the best paths generated by each algorithm. IAGOA produces a smooth, collision-free trajectory navigating the narrow corridor between the upper and lower obstacle clusters with minimal curvature. GOA and AGOA generate visually similar but slightly longer paths, consistent with their higher mean costs. GWO's path exhibits noticeable oscillations near obstacles, reflecting its tendency toward suboptimal local refinement. The PSO and GA paths frequently intersect obstacles, corroborating their high worst-

case costs and rendering them unreliable for safety-critical applications.

Fig. 3. Best Paths Generated by Each Algorithm



E. Statistical Validation

Wilcoxon Signed-Rank Tests: Table II presents pairwise comparisons between IAGOA and each competitor. IAGOA achieves statistical significance ($p < 0.05$) against all algorithms, including GWO ($p = 0.00926$) and AGOA ($p = 6 \times 10^{-5}$). The p-value of 0 against GOA, PSO, and GA indicates unequivocal superiority, with IAGOA outperforming these algorithms in every single run.

Table II. Wilcoxon Signed-Rank Tests (IAGOA vs. Competitors)

Comparison	W-statistic	p-value	Significant ($\alpha = 0.05$)
IAGOA vs GOA	20	0	YES
IAGOA vs AGOA	57	6×10^{-5}	YES
IAGOA vs PSO	0	0	YES
IAGOA vs GWO	119	0.00926	YES
IAGOA vs GA	0	0	YES

Friedman Test: The global rank-order test yields a chi-square statistic of $\chi^2 = 97.09$ ($p = 0$), rejecting the null hypothesis of equal algorithm performance. IAGOA secures the lowest mean rank, confirming it as the statistically superior optimizer among all six algorithms tested.

V. CONCLUSION AND FUTURE WORK

This paper proposed IAGOA, integrating five synergistic mechanisms that fundamentally address premature convergence and instability in canonical GOA. Extensive experiments demonstrate IAGOA's superior performance: a mean cost of 16.27 with standard deviation of 0.28, representing 23.3× stability improvement over GWO. The best-worst range of 1.43 units confirms near-deterministic consistency critical for autonomous navigation. Wilcoxon tests confirm statistical significance ($p < 0.05$) against all five competitors, and the Friedman test ($\chi^2 = 97.09, p = 0$) ranks IAGOA first. Future work will extend IAGOA to dynamic environments, multi-robot coordination, and real-world robotic validation.

REFERENCES

[1] M. N. Ab Wahab, S. Nefti-Meziani, and A. Atyabi, "A comparative review on mobile robot path planning: Classical or meta-heuristic methods?," *Annual Reviews in Control,* vol. 50, pp. 233-252, 2020.

[2] B. Patle, A. Pandey, D. Parhi, and A. Jagadeesh, "A review: On path planning strategies for navigation of mobile robot," *Defence Technology,* vol. 15, no. 4, pp. 582-606, 2019.

[3] J. Canny, *The complexity of robot motion planning*. MIT press, 1988.

[4] I. Noreen, A. Khan, and Z. Habib, "Optimal path planning using RRT* based approaches: a survey and future directions," *International Journal of Advanced Computer Science and Applications,* vol. 7, no. 11, pp. 97-107, 2016.

[5] S. Mirjalili, "The ant lion optimizer," *Advances in engineering software,* vol. 83, pp. 80-98, 2015.

[6] D. Karaboga and B. Akay, "A comparative study of artificial bee colony algorithm," *Applied mathematics and computation,* vol. 214, no. 1, pp. 108-132, 2009.

[7] S. Saremi, S. Mirjalili, and A. Lewis, "Grasshopper optimisation algorithm: theory and application," *Advances in Engineering Software,* vol. 105, pp. 30-47, 2017.

[8] Y. Meraihi, A. B. Gabis, S. Mirjalili, and A. Ramdane-Cherif, "Grasshopper optimization algorithm: theory, variants, and applications," *Ieee Access,* vol. 9, pp. 50001-50024, 2021.

[9] A. Gasparetto and V. Zanotto, "A new method for smooth trajectory planning of robot manipulators," *Mechanism and machine theory,* vol. 42, no. 4, pp. 455-471, 2007.

[10] M. Davoodi, F. Panahi, A. Mohades, and S. N. Hashemi, "Clear and smooth path planning," *Applied Soft Computing,* vol. 32, pp. 568-579, 2015.

[11] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-international conference on neural networks*, 1995, vol. 4: IEEE, pp. 1942-1948.

[12] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in engineering software,* vol. 69, pp. 46-61, 2014.

[13] D. E. Goldberg, "others, Genetic algorithms in search, optimization, and machine learning, vol. 412," ed: Addison-wesley Reading Menlo Park, 1989.

Yılan Oyunu Oynayan Bir Derin Pekiştirmeli Öğrenme Ajanının Tasarımı ve Performans Analizi

Fatih EYMEN
Mekatronik Mühendisliği Bölümü
Bursa Teknik Üniversitesi
Bursa, Türkiye
20406601056@ogr.btu.edu.tr

Dr. Seyitali İLYAS
Elektrik ve Elektronik Mühendisliği Bölümü
KIRGIZİSTAN-TÜRKİYE MANAS ÜNİVERSİTESİ
seyitali.ilyas@manas.edu.kg

Dr. Öğr. Üyesi İlhan TUNÇ
Mekatronik Mühendisliği Bölümü
Bursa Teknik Üniversitesi
Bursa, Türkiye
ilhan.tunc@btu.edu.tr

***Özet*—Pekiştirmeli öğrenme çevre ile etkileşimli ortamlarda deneme-yanılmaya dayalı karar verme yeteneği sayesinde otonom ajanlı sistemlerin geliştirilmesinde önemli bir araştırma konusudur. Özellikle durum uzayının karmaşık, büyük ve kararların ardışık alınmasının gerektiği durumlarda derin öğrenmeli derin pekiştirmeli öğrenme yöntemleri etkili çözümler sunmaktadır. Bu çalışmada ardışık karar verme ve büyük alan yönetimi gerektiren problemlere örnek teşkil eden Yılan (Snake) oyunu için Evrişimli Sinir Ağları (CNN) tabanlı iki farklı ajan geliştirilmiş ve performansları karşılaştırılma amacıyla analiz edilmiştir. Uzaysal algı ve dinamik ödül şekillendirme yaklaşımıyla birlikte On-Policy (Deep SARSA) ve Off-Policy (Deep Q-Learning) algoritmaları aynı çevresel değişkenler ve parametreler altında eğitilmiştir. İlk 2000 adet oyun üzerinden gerçekleştirilen deneysel sonuçlar, Deep SARSA ajanının Deep Q-Learning ajanına kıyasla daha yüksek bir ortalama skor ivmesi yakaladığını göstermektedir. Ayrıca, eğitim verileri incelendiğinde, On-Policy yaklaşımın model hatasının ani sıçramalarını engellediği ve ödül artışının daha az salınım sergileyerek hedefe daha sağlam bir şekilde yakınsadığını göstermiştir. Algoritmaların sadece elde ettikleri maksimum skor dışında sistem kararlılığı ve öğrenme dinamikleri üzerinden de değerlendirilmesi gerektiğini ortaya koyarak otonom karar verme sistemlerinin tasarımına yönelik kapsamlı bir çerçeve sunmaktadır.**

Anahtar Kelimeler: Pekiştirmeli öğrenme, derin pekiştirmeli öğrenme, Deep Q-Learning, Deep SARSA, On-Policy, oyun yapay zekâsı, otonom ajanlar.

Abstract — Reinforcement learning is a important research area in the development of autonomous systems because of its ability to make decisions based on trial-and-error in interactive autonomous environment. Deep reinforcement learning methods supported by deep neural networks offer effective solutions where the state space is so complex, environmental changes are fast and decisions are all sequential. Two agents trained on Convolutional Neural Networks (CNN) were designed to solve problems for the Snake game which problems requiring sequential decision-making and dynamic space management. Their performances were comparatively analyzed. Utilizing the proposed ego-centric space perception and dynamic reward shaping approaches, On-Policy (Deep SARSA) and Off-Policy (Deep Q-Learning) algorithms were trained under same environmental conditions. Experimental results conducted over the first 2000 training episodes demonstrate that the Deep SARSA agent achieved a higher average score momentum compared to the Deep Q-Learning agent. The analysis of training metrics revealed that the On-Policy approach prevented sudden spikes in model loss and exhibited reduced oscillation in reward maximization, converging to the optimal policy in a significantly more robust manner. This study demonstrates that RL algorithms should be evaluated not solely based on the maximum score achieved, but also on system stability and learning dynamics, providing a comprehensive framework for the design of autonomous decision-making systems.

Index Terms — Reinforcement learning, deep reinforcement learning, Deep Q-Learning, Deep SARSA, On-Policy, game AI, autonomous agents.

I. GIRIŞ

Ajanların büyük ve karmaşık ortamlarda bağımsız karar verebilme yeteneği, günümüz yapay zekâ çalışmalarının önemli uygulama alanlarından biridir **[1]**. Pekiştirmeli Öğrenme (RL), bu alanda sık kullanılan bir yöntemdir ve ajanın oyun ortamıyla girdiği deneme-yanılma etkileşimleri sonucunda en iyi davranış politikasını (policy) oluşturmasını sağlar **[2]**. Ajan, her seferinde aldığı kararlara göre çevreden gelen ödül veya ceza sinyalini takip ederek anlık kazançlardan ziyade uzun vadeli başarıyı en yüksek seviyede tutacak matematiksel bir model kurmayı hedefler **[3]**.

Algoritmaların sınırlarını sınamak Yılan (Snake) oyunu araştırmacıların ilgisini çekmektedir **[4]**. Yılan oyununun temel kuralları basit olmasına rağmen oyun ortamı ve kendi dinamikleri nedeniyle oyun oynayan ajan için kendine özgü zorluklar bulunmaktadır. Yılanın yemi her aldığında gövdesinin uzaması ve buna bağlı olarak hareket alanının giderek daralması, karar verme sürecini değişken ve doğrusal olmayan bir probleme dönüştürmektedir. Pekiştirmeli öğrenme algoritmalarının bu dinamik koşullar altındaki performansını ölçmek amacıyla, Pygame **[5]** kütüphanesi kullanılarak probleme göre değiştirilebilir bir simülasyon ortamı oluşturulmuştur. Pygame'in sağladığı ızgara tabanlı oyun alanı, ajanın çevreyle olan etkileşimini ayrık adımlara bölerek her bir durumun (state) algoritmaya düzenli ve kayıpsız biçimde aktarılmasını sağlamıştır.

Literatür incelemelerinden sonra Yılan oyunu ve benzeri ajan navigasyonu problemlerinde araştırmacıların genellikle tek bir algoritmaya ve daha küçük oyun ortamlarına odaklandığı görülmektedir. Literatürdeki güncel çalışmalardan birinde, klasik Snake oyununu çözmek için yalnızca Deep Q-Network (DQN) algoritması kullanılmış ve ajan 20 x 20 boyutlarındaki görece dar bir ortamda eğitilerek maksimum 60 bandında skorlara ulaşabilmiştir **[6].** Bunun yanında, yalnızca ajanın elde ettiği en yüksek puanı temel alan mimariler üzerine yoğunlaşıldığı görülmektedir **[7].** Izgara boyutu arttıkça oyun alanının büyümesi, ajanın öğrenme

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

sürecini ve yerel optimumdan kaçma kabiliyetini zorlaştırmaktadır. Ajanlar literatürdeki diğer çalışmalara göre çok daha geniş ve yönetilmesi zor olan 32 x 32 boyutlarındaki bir eğitim alanında teste tabi tutulmuştur. Bu büyüklükteki oyun alanına rağmen geliştirilen model, literatürdeki skor bantlarına yakın sonuçlara ulaşabilmiştir.

Literatürdeki ışın dökümü (raycasting) gibi doğrusal algılama sensörlerinin oluşturduğu kör noktaları aşmak için ajana Evrişimli Sinir Ağları (CNN) aracılığıyla doğrudan karmaşık algılama yeteneği kazandırılmıştır **[8]**, **[9]**. Ajanın durum temsilindeki karmaşıklığı sadeleştirmek için 32 x 32'lik oyun alanı, her adımda ajanın yönelimine göre döndürülerek öğrenme sürecindeki yön bağımlılığı azaltılmıştır. Eğitimlerin amacı, tekil model yaklaşımından ziyade kurulan mimariler üzerinden On-Policy Deep SARSA ve Off-Policy Deep Q-Learning algoritmalarının performanslarını aynı parametreler altında karşılaştırmalı olarak incelemektir. Analizlerde algoritmalar yalnızca ulaştıkları en yüksek skor üzerinden karşılaştırılmamış; öğrenme hızları, hata azaltma eğilimleri ve stabilite verileri üzerinden de değerlendirilmiştir.

32x32 boyutlarındaki geniş oyun alanında ajanın oyun alanını algılayabilmesi için 11x11 boyutunda iki kanallı bir alan tarayıcı tasarlanmış ve ardışık evrişimli katmanları olan CNN mimarisi kurulmuştur. Ajanın oyun alanı karmaşıklığını azaltmak adına bakış açısına göre (yönelimine bağlı) referans sistemi entegre edilmiştir. Yılan uzadıkça yılanı kendinden korumak için ve eğitimin sürdürülebilir olabilmesi adına "Uzun Yılan Toleransı" stratejisini içeren dinamik bir ödül şekillendirme (reward shaping) mekanizması kullanılmıştır. Geliştirilen bu bütünsel mimari üzerinde Deep Q-Learning ve Deep SARSA algoritmaları GPU ivmelendirmesiyle kararlı halde çalışırken ajanların öğrenmelerinin kapatıldığı bağımsız bir sömürü (exploitation) testi gerçekleştirilerek elde edilen modeller hem davranış bakımından hem de performans bakımından analiz edilmiştir.

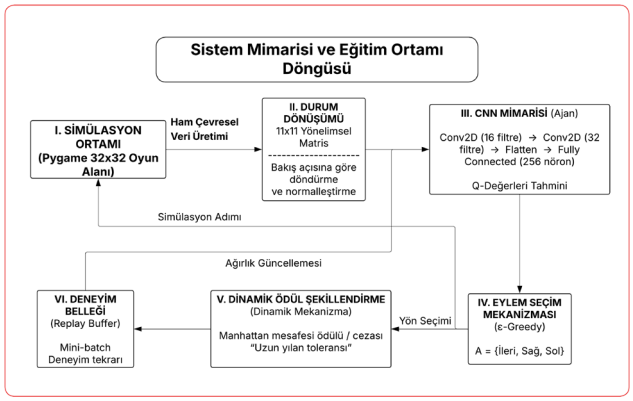
Çalışmanın devam eden bölümleri şu şekilde yapılandırılmıştır. İkinci bölümde önerilen yöntem ayrıntılı olarak açıklanmış, üçüncü bölümde gerçekleştirilen benzetim çalışmaları ve bu çalışmalardan elde edilen bulgular değerlendirilmiştir. Dördüncü bölümde ise çalışmanın genel sonuçları sunularak elde edilen çıktılar yorumlanmıştır.

II. Sistem Mimarisi ve Yöntem

Eğitimlerde derin pekiştirmeli öğrenme tabanlı ajan eğitimi için tasarlanan mimari (sinir ağı modelleri) ve oyun alanı (simülasyon ortamı) kapalı çevrim etkileşime sahiptir. Tasarlanan sistem mimarisi, yüksek boyutlu alandan öznitelik çıkarımı, politika optimizasyonu ve ödül şekillendirme mekanizmalarının entegrasyonundan oluşur.

A. Sistem Mimarisi ve Çevresel Döngü

Eğitimin kontrol sisteminin veri akış yönleri Şekil 1'de modellenmiştir. Simülasyon ortamı 32x32 boyutlarında ızgara (grid) matris olarak seçilmiştir. Şekil 1'de şematize edilen kapalı çevrim eğitim döngüsü ilk adımda gövde koordinatları, duvar sınırları ve yemin anlık konumu olan ham harita verilerini üretir. İkinci adımda ajan ham harita üzerinden yön algısı sağlayan 11x11 boyutlarında yönelim tarayıcı bir matrisle arama yapar ve veriyi sadeleştirir. Üçüncü adımda geliştirilen Evrişimli Sinir Ağı (CNN) mimarisi her hareketi ardışık evrişim katmanlarından geçirerek eylem uzayındaki her bir hareket için tahmini Q değerlerini üretir. Dördüncü adımda tahmin edilen değerler doğrultusunda yönler(sağ, sol, ileri) o anki keşif/sömürü (ε-Greedy) politikasına uygun yön seçilir. "Simülasyon Adımı" olarak doğrudan simülasyon ortamına adımlar iletilir. Beşinci adımda oyun alanı değişirken eylemin durumuna ve yılanın uzunluğuna (Manhattan mesafesi ve Uzun Yılan Toleransı) bağlı olarak bir ödül/ceza sinyali oluşur. Altıncı adımda elde edilen deneyim verileri(durum, eylem, ödül, yeni durum) optimizasyon amacıyla deneyim belleğine (Replay Buffer) aktarılır. Mini-batch yöntemiyle hafıza havuzundan çekilen rastgele deneyimler CNN mimarisinin (üçüncü adım) ağırlıklarını güncelleyerek öğrenme sürecini tamamlar.



Şekil 1. Önerilen derin pekiştirmeli öğrenme sistem mimarisi ve eğitim ortamı döngüsü.

Bu çalışmada eğitilen RL ajanının karar süreci durumlara, eylemlere ve ödül arasındaki bağlamların olduğu Markov Karar Süreci (MDP) tarafından yapılmıştır **[2]**. Ajan eğitim performansını iyileştirmek için bellek yönetimi, uzaysal algı ve dinamik ödül stratejileriyle eğitimler gerçekleştirilmiştir.

B. Uzaysal Algı ve CNN Mimarisi

Ajanın eğitim ortamını algılayabilmesi için normalde kullanılan 1 boyutlu mesafe vektörleri yerine, 11x11 boyutlarında 2 boyutlu bir kare matris kullanılmıştır.

- **İlk Kanal:** Duvarlar ve yılanın kendi gövdesi 1 değeri ile gösterilmiştir.
- **İkinci Kanal:** Yem ise yendikten sonra rastgele bir nokta olarak tanımlanmaktadır.

Modelin yakınsaması hızını çoğaltmak ve durum karmaşıklığını en aza indirebilmek için yönelimsel algıya dayalı bir referans sistemi uygulanmıştır. Ajanın her hamlesini 11x11'lik matris mevcut bakış açısına göre döndürerek işlemektedir. Ajan bu sayede hareket alanı dört yönden üç bağıl yöne (ileri, sağ, sol) indirmiş olduk, böylece kendi üzerine dönerek (180 derece) intihar etmesi engellendi. Kurduğumuz mimariyle verilerin işlenmesi için ardışık iki Evrişimli Katman (Conv2D) kullanılmıştır **[7]**. İlk katman 16 filtre, ikinci katman 32 filtre ile özellik çıkarımı yaparken elde edilen öznitelikler 256 nöronlu bir tam bağlantılı(fully connected) katmandan geçirilerek eylemlere ait Q değerlerini üretmektedir.

C. Karar Verme Algoritmaları: On-Policy ve Off-Policy Yaklaşımları

Kullanılan mimarinin karar verme kabiliyetini gözlemlemek için iki farklı pekiştirmeli öğrenme algoritması teste tabi tutulmuştur.

- **Deep Q-Learning (Off-Policy):** Yöntemde amaçlanan durum ajanın mevcut koşuldaki keşif politikasından bağımsızca sonraki adımda en

yüksek ödülü sağlayacak hareketleri yapmasıdır **[3]**, **[10]**. Ağırlık değişimleriyle her zaman en yüksek Q değerine ulaşmaya çalışarak agresifçe çalışır.

$$Q(s,a) \leftarrow Q(s,a) + \alpha[R + \gamma \backslash max_{a'} Q(s',a') - Q(s,a)]$$

- **Deep SARSA (On-Policy):** Deep Q-Learning'in aksine Deep SARSA hedefe giden yolunu hesaplarken varsayıma göre hesaplama yerine ajan politikasının (ε -greedy) gerçekten seçeceği bir sonraki hamleyi (a') dikkate alır **[11]**. Deep SARSA yönteminde ajan daha temkinli bir tutum sergiler.

$$Q(s,a) \leftarrow Q(s,a) + \alpha[R + \gamma Q(s',a') - Q(s,a)]$$

Modellerin yüksek boyutlara ulaşan verilerle stabil şekilde eğitilebilmesi için Deneyim Tekrarı (Experience Replay) kullanılarak deneyimler kayıt altına alınmıştır **[12]**. Ajanın eğitimi sırasındaki tecrübeleri geniş bir havuzda (replay buffer) saklanmıştır. Eğitimde havuzdan rastgele alınan mini-gruplar (batch) Evrişimli Sinir Ağının ağırlıklarını güncelleyerek öğrenme sürecini hızlandırmıştır. Bu yöntemler eğitimde ardışık durumlar arasındaki benzer ilişkileri azaltıp ağın aşırı uyum (overfitting) yapmasını engellemekte ve matris çarpımlarının tam kapasiteyle kullanılmasını sağlamaktadır.

D. Dinamik Ödül Şekillendirme

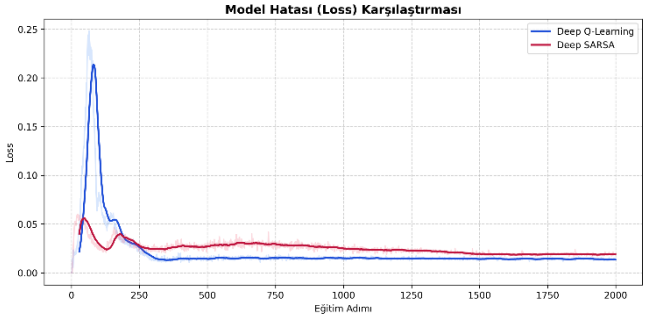
Ajan hedefe ulaşsın ama duvarlara da çarpmasın diye yem için +10 ve duvara çarpmalar için -10 puan alacağı ödüllendirme ve cezalandırma sistemi eklenmiştir **[6]**. Manhattan mesafesi optimizasyonuyla yeme yaklaşan adımlar +0.1 ile ödüllendirilirken uzaklaşan adımlar -0.2 ile cezaya tabi olmaktadır. Yılan uzadıkça büyüyen gövdesi yeme giden en kısa yolu tıkayabilmektedir ve bu yüzden eğitimin ileri adımlarına ulaşmada problemler yaşamaktadır. Ajanın mesafe cezasından kurtulmak için dar alanlara girip kendine vurmasını engellemek amacıyla "Uzun Yılan Toleransı" uygulanılmıştır. Ajan boyu belirlenen büyüklüğü aştığında yoldan sapma cezası kaldırılmaktadır. Bu ödül yapısı sayesinde ajan kısayken hedefe agresif bir şekilde yönelirken yılan uzadığında kendi kuyruğu etrafından geçebilsin diye cezalandırma sistemi yoldan bağımsızlaştırılarak yılana hareket özgürlüğü verilmiştir.

III. BENZETİM ÇALIŞMALARI

Geliştirilen uzaysal algılı CNN ve dinamik ödül şekillendirme eğitim modeli üzerinde kurgulanan On-Policy (Deep SARSA) **[9]** ve Off-Policy (Deep Q-Learning) **[3]** ajanların eğitim performansları 32x32 boyutlarında eğitim ortamında eğitilerek kıyaslamaya uygun hale getirildi. Yüksek boyutlu matris çarpımları çok fazla olacağı ve uzun süreceği için hesaplama yükü tolere edebilmek amacıyla eğitimler GPU üzerinden gerçekleştirilmiştir. Ajanların öğrenme süreçlerinin adil bir şekilde değerlendirilebilmesi kıyaslamanın adil olacağı şekilde her iki algoritmanın da ilk 2000 oyun boyunca elde ettiği skor, model hatası ve toplam ödül verileri kayıt altına alınarak karşılaştırılmıştır.

A. Bireysel Eğitim Dinamikleri ve Yakınsama (Convergence)

Eğitimin başlangıcında (keşif oranı $\varepsilon > 0.8$) her iki ajan da ortamın ağırlıklı olarak rastgele adımlarla öğrenilmesine ve düşük hayatta kalma sürelerine sebep olmuştur. ε değeri azaldıkça sömürü evresinde(exploitation) Deneyim Tekrarı (Experience Replay) **[12]** mekanizmasının da etkisiyle Q-değerleri stabil bir hale gelmişti

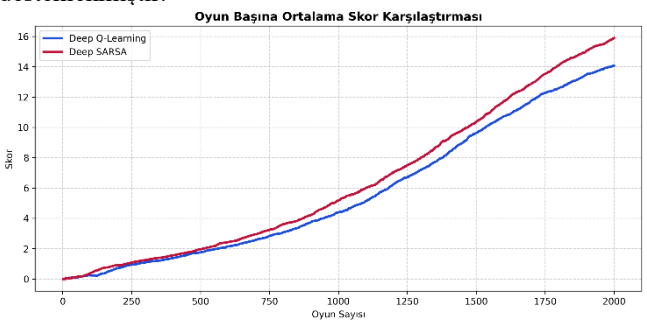


Şekil 2. Ajanların ilk 2000 oyundaki model hatalarının karşılaştırılması

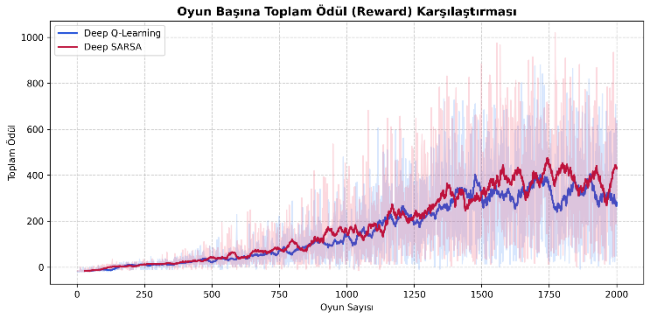
Şekil 2'deki Model Hatası (Loss) grafiklerine bakıldığında karar verme sistemlerinin matematiksel altyapılarından kaynaklanan bazı belirgin farklılıklar görülmüştür. Deep Q-Learning modelinin eğitim evresinin başında yüksek hata sıçramaları yaşanmıştır. Deep SARSA modelinde ise hata grafiği çok daha stabil kalabilmiştir. Erken sönümlemeyle model için daha stabil ve daha pürüzsüz bir yakınsama (convergence) elde edildiğine işaret edilmektedir.

B. Karşılaştırmalı Performans ve Stabilite Analizi

Politika yaklaşımlarının oyun performansına yaptığı etkiyi incelemek için ortalama skor ve toplam ödül verileri aynı eksen üzerinde bir araya getirilerek nicel verilerle desteklenmiştir.



Şekil 3. Deep Q-Learning ve Deep SARSA ajanlarının ortalama skor bazında karşılaştırmalı analizi



Şekil 4. Deep Q-Learning ve Deep SARSA ajanlarının toplam ödül bazında karşılaştırılması

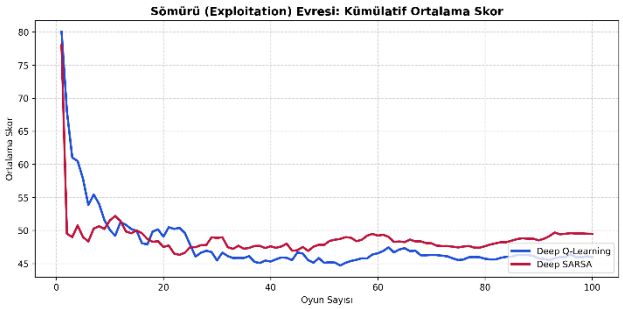
TABLO I. *İLK 2000 OYUN SONUNDA ALGORITMA PERFORMANS ANALIZI*

Performans Metriği	**İlk 2000 Oyun Sonunda Algoritma Performans Analizi**	
	Deep Q-Learning (Off-Policy)	***Deep SARSA (On-Policy)***
Ulaşılan En Yüksek Skor	62	54
2000. Oyunda Ortalama Skor	14.12	15.85
Maksimum Model Hatası (Loss Zirvesi)	0.245	0.059
Öğrenme Kararlılığı(Robustness)	Düşük (Agresif)	Yüksek (Temkinli)

Şekil 3'teki ortalama skor grafiğine, Şekil 4'teki toplam ödül grafiğine ve Tablo I'deki verilere bakılarak ilk 500 oyunda nerdeyse aynı performansta ilerleyen modellerin 1000. oyundan sonra aradaki farkı açmaya başladıkları gözlemlenmiştir. Off-Policy algoritma (Deep Q-Learning) ödüle giderken, agresifliği **[8]** zaman zaman ani skor artışlarına yol açtığı gibi oyun başına toplam ödül grafiğindeyse salınımların oluşmasına yol açmıştır. On-Policy yaklaşım (Deep SARSA) rakip ajana göre daha temkinli bir politika izlemiştir **[9]**. Bu sayede 2000. oyunda daha yüksek ortalama skora (15.85) ulaşmıştır.

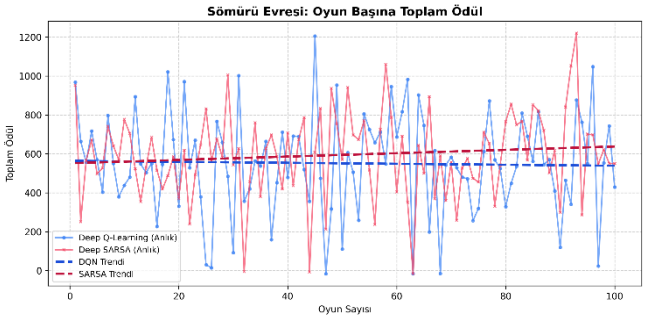
C. Sömürü (Exploitation) Evresinde Davranışsal Analiz

Ajanların eğitim algoritmalarının en iyi performanslarını gözlemleyebilmek için keşif oranının sıfırlandığı ($\varepsilon = 0$) ve ağırlıkların değişimlerinin kapatılarak sadece öğrenilmiş politikaların oynandığı 100 oyunluk sömürü testi yapıldı. Testler sonucunda bulunan kümülatif ortalama skorlara bakılınca (Şekil 5) Deep SARSA ajanın 20. oyundan itibaren stabilitesini sağladığı ve 100 oyunun sonunda yaklaşık 50 skor civarına yerleşerek daha iyi performans sağladığı görülmüştür. Deep Q-Learning (DQN) ajanı ise aynı koşullarda 46 skor civarlarında kalıp daha düşük kümülatif ortalama elde edebilmiştir.



Şekil 5. Deep Q-Learning ve Deep SARSA algoritmalarının sömürü evresindeki (100 oyun) kümülatif ortalama skor performansı

Toplam ödül grafiğine (Şekil 6)bakacak olursak otonom kararların doğası gereği anlık oyunlarda yüksek salınımlar görülüyor olsa da modellerin trend çizgileri birbirinden ayrılmaktadır. DQN ajanının toplam ödül trendi test boyunca yatay eğilim sergilerken SARSA ajanının trend çizgisi belirgin bir şekilde artmıştır.



Şekil 6. Sömürü evresinde elde edilen oyun başına toplam ödül dağılımı ve algoritmaların eğilim (trend) çizgileri

Off-Policy yapısında olan DQN ajanı ilerleyen her oyunda açgözlü bir strateji izlediği için ödülüne giden en kısa, tehlikeli yolları seçmiş ve ani çarpışmalarla trendin düşmesine neden olmuştur. On-Policy yapısındaki SARSA ajanı uzayındaki riskleri hesaba katarak tehlikeli hamlelerden kaçınarak yeme ulaşmak için yolunu uzatıp kendi kuyruğunun etrafından dolanan manevraları tercih etmiştir. SARSA'nın yükselen ödül trendi temkinli davranan ajanımızın hayatta kalma süresini artırdığı ve uzun vadede ajanın stabilizasyonunu olabilecek en yüksek seviyeye çıkardığını ortaya koymuştur.

D. Eğitim Hiperparametreleri

Modeller eğitim süreçlerindeyken çevresel şartların sabit tutulması ve sonuçların tekrarlanabilirliği adına Tablo II'de belirtilen hiperparametreler kullanılmıştır.

TABLO II. EĞITIM HIPERPARAMETRELERI

Hiperparametre	Parametre Detayları	
	Açıklama	***Değer***
Öğrenme Hızı(α)	Ağ ağırlıklarının güncellenme katsayısı (Learning Rate)	0.0003
İndirim Faktörü(γ)	Gelecekteki ödüllerin anlık kararlara etkisi (Discount Factor)	0.99
Hafıza Kapasitesi	Deneyim havuzunun maksimum satır boyutu	200000
Eğitim Paketi (Batch Size)	Replay Buffer'dan tek seferde çekilen deneyim sayısı	256
Keşif Azalma Oranı (ε)	Rastgele keşif ihtimalinin sönümlenme çarpanı	0.9975
Hedef Ağ Güncelleme (τ)	Target ağın ana ağdan kopyalanma hızı (Soft-Update)	0.005

IV. SONUÇ VE GELECEK ÇALIŞMALAR

Karmaşık dinamiklere ve geniş bir durum uzayına (32x32) sahip Yılan (Snake) oyunu probleminde otonom bir ajanın karar verme kabiliyeti Evrişimli Sinir Ağları (CNN) tabanlı derin pekiştirmeli öğrenme modelleriyle çalışmanın konusu oldu. Geleneksel doğrusal algılayıcıların sahip olduğu kör alanlar ajanın çevresini bir matris tarayıcıyla algıladığı mimariyle aşılarak "Uzun Yılan Toleransı" ve dinamik ödül şekillendirme yapısıyla ajanı kör alanlara takılmaktan uzak tutabilmiştir.

CNN mimarisi ve aynı ortamsal kısıtlar altında eğitilen Deep Q-Learning (Off-Policy) ve Deep SARSA (On-Policy) algoritmalarının performansları karşılaştırılarak analiz edilmiştir. Deep SARSA ajanının bir sonraki hedefe giden riskleri de politikasına dahil eden temkinli yapısıyla model hatası (loss) iniş çıkışlarını minimize ettiğini ve daha stabil bir öğrenme performansı göstermiştir. Öğrenmenin kapatıldığı 100 oyunluk sömürü (exploitation) testindeki veriler bu durumun Deep Q-Learning ajanının agresif stratejisi nedeniyle hata yapmaya yatkınlık gösterip yatay bir ödül trendiyle 46 skor bandında kalmasına neden olmuştur. Deep SARSA ajanı daha uzun rotalarla pozitif bir ödül trendi yakalamış ve 50 skor civarında stabil bir şekilde yer almıştır. Ajan eğitimlerinde algoritma seçiminin sadece ulaşılacak maksimum skoru değil, sistemin çevresel risklere karşı duruşunu ve değişken ortamlara olan adapte olma kabiliyetini de ortaya koymuştur.

Gelecek çalışmalarda oyun ortamının karmaşıklığını artırmak ve politikaların tutumlarını dinamik sorunlara karşı daha iyi ölçebilmek amacıyla oyun alanına rastgele veya belirli alanlarda hareket eden dinamik engellerin entegre edilmesi düşünülmektedir. Bu sayede ajanın yalnızca statik duvarlara ve kendi gövdesinden ziyade sürekli değişen çevresel tehditlere karşı da anlık karar alma kabiliyeti test edilecektir. Geliştirilen CNN mimarisinin ve üzerine inşa edilecek dinamik engellerden kaçınma probleminin ilerleyen aşamalarda üretim tesislerinde veya insan bulunduran alanlarda görev yapan robotların hareketli hedeflere karşı yürüttüğü gerçek zamanlı endüstriyel navigasyon ve iş sağlığı problemlerine temel oluşturması öngörülmektedir.

TEŞEKKÜR

Bu çalışmada kullanılan pekiştirmeli öğrenme algoritmalarının teorik açısından, model eğitim stratejilerinden ve teknik altyapı konusundaki değerli rehberliği ve sağladığı bilgi birikimi için Sayın Dr. İlhan Tunç'a içtenlikle teşekkür ederim. Tüm destekleri için minnettarım.

REFERENCES

[1] *S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Pearson, 2020.*

[2] *R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.*

[3] *V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529–533, Feb. 2015.*

[4] *G. N. Yannakakis and J. Togelius, Artificial Intelligence and Games. Springer, 2018.*

[5] *P. Shinners, "Pygame," pygame.org, 2011. [Online]. Available: https://www.pygame.org.*

[6] *F. C. et al., "A Deep Q-Learning based approach applied to the Snake game," 2021 29th Mediterranean Conference on Control and Automation (MED), 2021.*

[7] *"Deep Q-Snake: An Intelligent Agent Mastering the Snake Game with Deep Reinforcement Learning," IEEE Conference Publication, 2024.*

[8] *I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.*

[9] *F. Chollet, Deep Learning with Python. Shelter Island, NY, USA: Manning Publications, 2017.*

[10] *C. J. C. H. Watkins and P. Dayan, "Q-learning," Machine Learning, vol. 8, pp. 279-292, 1992.*

[11] *G. A. Rummery and M. Niranjan, "On-line Q-learning using connectionist systems," Cambridge University Engineering Department, Tech. Rep. CUED/F-INFENG/TR 166, 1994.*

[12] *L. J. Lin, "Self-improving reactive agents based on reinforcement learning, planning and teaching," Machine Learning, vol. 8, pp. 293-321, 1992.*

Pekiştirmeli Öğrenme Tabanlı Trafik Işığı Kontrolü: Ata Bulvarı-Fethiye Kavşağı Simülasyonu

Reinforcement Learning-Based Traffic Signal Control: Simulation of the Ata Boulevard-Fethiye Intersection

Şeyma DURMUŞ
BURSA TEKNİK ÜNİVERSİTESİ
Mekatronik Mühendisliği
Bursa, Türkiye
22406601021@ogr.btu.edu.tr

Dilara ÇAYLAKOĞLU
BURSA TEKNİK ÜNİVERSİTESİ
Mekatronik Mühendisliği
Bursa, Türkiye
23406601062@ogr.btu.edu.tr

Zişan BAŞARAN
BURSA TEKNİK ÜNİVERSİTESİ
Mekatronik Mühendisliği
Bursa, Türkiye
21406601037@ogr.btu.edu.tr

Resul İLİMDAR
BURSA TEKNİK ÜNİVERSİTESİ
Mekatronik Mühendisliği
Bursa,Türkiye
23406601056@ogr.btu.edu.tr

Dr. Seyitali İLYAS
KIRGIZİSTAN-TÜRKİYE MANAS ÜNİVERSİTESİ
Elektrik-Elektronik Mühendisliği
Bişkek, Kırgızistan
seyitali.ilyas@manas.edu.kg

Dr. Öğr. Üyesi İlhan TUNÇ
BURSA TEKNİK ÜNİVERSİTESİ
Mekatronik Mühendisliği
Bursa, Türkiye
ilhan.tunc@btu.edu.tr

***Özet*— Artan araç yoğunluğu, şehir altyapısının mevcut kapasitesini aşarak kent içi ulaşımda trafik sıkışıklığına neden olmaktadır. Bu durum, bekleme sürelerinin uzaması ve karbon dioksit (CO_2) emisyonlarının artması gibi önemli problemlere yol açmaktadır. Bu çalışmada, trafik sıkışıklığının en yoğun olarak gözlendiği noktalar olan kavşaklara odaklanılmış ve kavşaklardaki trafik ışıklarının kontrolü yoluyla trafik sıkışıklığının azaltılması amaçlanmıştır. Söz konusu probleme çözüm geliştirmek amacıyla, Bursa ili Nilüfer ilçesinde yer alan ve yüksek trafik hacmine sahip Ata Bulvarı–Fethiye kavşağı çalışma alanı olarak seçilmiştir. Bu kapsamda, SUMO (Simulation of Urban Mobility) benzetim ortamında Ata Bulvarı–Fethiye kavşağı modellenmiş ve kavşak kontrolü için pekiştirmeli öğrenme tabanlı bir trafik ışığı kontrolü yaklaşımı geliştirilmiştir. Geliştirilen yöntemde, araç yoğunluğu ve bekleme süreleri gibi trafik parametreleri dikkate alınarak pekiştirmeli öğrenme temelli bir trafik ışığı kontrol ajanı eğitilmiştir. Eğitilen ajan test edilmiş ve elde edilen benzetim sonuçları, önerilen yaklaşımın geleneksel trafik ışığı kontrol yöntemlerine kıyasla araç bekleme sürelerinin, kuyruk uzunluklarının ve CO_2 emisyonlarının azaltılmasında etkili olduğunu göstermiştir. Sonuç olarak, pekiştirmeli öğrenme tabanlı uyarlamalı kontrol yöntemlerinin akıllı ulaşım sistemleri kapsamında kent içi trafik yönetimi açısından önemli bir potansiyel taşıdığı değerlendirilmiştir.**

Anahtar Kelimeler — pekiştirmeli öğrenme, trafik ışığı kontrolü, kentsel trafik yönetimi, kavşak optimizasyonu, SUMO

Abstract— Increasing vehicle density exceeds the existing capacity of urban infrastructure and causes traffic congestion in urban transportation. This situation leads to significant problems such as longer waiting times and increased carbon dioxide (CO_2) emissions. In this study, intersections, where traffic congestion is most commonly observed, are focused on, and it is aimed to reduce traffic congestion through the control of traffic signals at intersections. In order to develop a solution to this problem, the Ata Boulevard–Fethiye intersection, located in the Nilüfer district of Bursa and having high traffic volume, was selected as the study area. Within this scope, the Ata Boulevard–Fethiye intersection was modeled in the SUMO (Simulation of Urban Mobility) simulation environment, and a reinforcement learning-based traffic signal control approach was developed for intersection control. In the proposed method, a reinforcement learning-based traffic signal control agent was trained by considering traffic parameters such as vehicle density and waiting times. The trained agent was tested, and the obtained simulation results showed that the proposed approach is effective in reducing vehicle waiting times, queue lengths, and CO_2 emissions compared to conventional traffic signal control methods. As a result,

reinforcement learning-based adaptive control methods are considered to have significant potential for urban traffic management within the scope of intelligent transportation systems.

Keywords: reinforcement learning, traffic signal control, urban traffic management, intersection optimization, SUMO

I. GİRİŞ

Şehir içi ulaşımda araç sayısının artması, özellikle sinyalize kavşaklarda bekleme süresi, kuyruk uzunluğu ve yakıt tüketimi gibi problemlerin daha belirgin hâle gelmesine neden olmaktadır. Bu nedenle trafik ışıklarının sabit süreli yapılar yerine anlık trafik durumuna göre uyarlanabilen yöntemlerle kontrol edilmesi önemli bir araştırma konusu hâline gelmiştir.

Sabit süreli trafik ışığı kontrol sistemleri, kavşak yönetiminde uzun yıllardır kullanılan temel yaklaşımlardan biridir. Ancak bu sistemlerde faz süreleri önceden belirlendiği için trafik yoğunluğunun gün içinde değiştiği durumlara yeterince uyum sağlanamamaktadır. Bu durum özellikle yoğun saatlerde bekleme süresinin, kuyruk uzunluğunun ve dur-kalk davranışının artmasına neden olmaktadır. Bu nedenle son yıllarda trafik ışığı kontrolünde, anlık trafik durumuna göre karar verebilen pekiştirmeli öğrenme tabanlı yöntemler öne çıkmaktadır.

Pekiştirmeli öğrenme tabanlı trafik ışığı kontrolünde ajan, kavşaktan aldığı durum bilgisine göre uygun faz kararını öğrenmeye çalışır. Liang ve arkadaşları, kavşağı hücrelere ayırarak araç konumlarını durum uzayına aktarmış ve derin pekiştirmeli öğrenme ile trafik ışığı çevrim sürelerini dinamik olarak belirlemiştir [1]. IntelliLight çalışmasında ise gerçek trafik verileri kullanılarak derin pekiştirmeli öğrenme tabanlı akıllı trafik ışığı kontrolü geliştirilmiş ve ajanın öğrendiği karar politikaları incelenmiştir [2].

Literatürdeki ilerleyen çalışmalarda faz seçimi, ödül tasarımı ve trafik akışının daha dengeli yönetilmesi üzerine odaklanılmıştır. PressLight çalışmasında maksimum basınç yaklaşımı pekiştirmeli öğrenme ile birleştirilerek arter yol ağlarında sinyal koordinasyonu hedeflenmiştir [3]. FRAP çalışmasında ise fazlar arasındaki rekabet ilişkisi dikkate alınarak hangi trafik hareketine öncelik verileceği öğrenilmiştir [4]. Bu çalışmalar, trafik ışığı kontrolünde yalnızca algoritmanın değil, faz yapısı ve ödül fonksiyonunun da model başarısı üzerinde etkili olduğunu göstermektedir.

Durum uzayı ve öğrenme yapısının performansa etkisi de literatürde önemli bir araştırma konusudur. Wu ve arkadaşları, derin pekiştirmeli öğrenme tabanlı trafik sinyal kontrolünde durum tanımı, ödül fonksiyonu ve öğrenme yönteminin etkisini karşılaştırmalı olarak incelemiştir [5]. Zhang ve arkadaşları, Deep Q-Learning yapısını bulanık mantık desteğiyle ele alarak sinyal kararlarının trafik yoğunluğuna göre daha esnek verilmesini amaçlamıştır [6]. Cai ve arkadaşları ise dueling double deep Q-network ve prioritized experience replay gibi yapıları kullanarak değişken trafik koşullarına karşı daha dayanıklı bir model önermiştir [7]. EP-D3QN çalışmasında da Double-Dueling DQN yapısı ile faz geçişleri ve yeşil sürelerinin daha dengeli ayarlanması hedeflenmiştir [8].

Pekiştirmeli öğrenme tabanlı yöntemlerin farklı senaryolarda karşılaştırılabilmesi için benchmark ve simülasyon ortamları da geliştirilmiştir. RESCO, trafik sinyal kontrol algoritmalarını ortak senaryo ve ölçütlerle değerlendirmek için önerilen bir benchmark yapısıdır [9]. CityFlow ise büyük ölçekli şehir trafik senaryolarında çok ajanlı pekiştirmeli öğrenme uygulamalarını destekleyen hızlı bir simülasyon ortamı sunmaktadır [10]. Çoklu kavşak yönetimi açısından CoLight çalışmasında graf dikkat ağları kullanılarak komşu kavşak bilgilerinin sinyal kararlarına dahil edilmesi sağlanmıştır [11]. Federated reinforcement learning yaklaşımında ise farklı kavşaklarda öğrenilen bilgilerin ortak bir modelde birleştirilmesi hedeflenmiştir [12].

Son dönem çalışmalarda yalnızca araç bekleme süresi değil, çevresel etkiler ve yaya hareketleri de dikkate alınmaktadır. EcoLight çalışmasında trafik ışığı kontrolü CO_2 emisyonu ve yakıt tüketimi gibi çevresel ölçütler üzerinden değerlendirilmiştir [13]. SUMO ve TraCI tabanlı yaya dostu trafik sinyal kontrolü çalışmasında ise araç akışı ile yaya geçiş talepleri birlikte ele alınmıştır [14]. Bursa özelinde yapılan Şehreküstü ve Gökdere çalışmasında da SUMO ortamında gerçek bölgeye dayalı derin Q-öğrenme tabanlı trafik ışığı kontrolü incelenmiştir [15].

Bu çalışmalar genel olarak, pekiştirmeli öğrenme tabanlı trafik ışığı kontrolünün bekleme süresi, kuyruk uzunluğu, gecikme ve çevresel etkiler açısından önemli bir potansiyel taşıdığını göstermektedir. Bu çalışmada ise Bursa ili Nilüfer ilçesinde yer alan Ata Bulvarı-Fethiye Kavşağı ele alınmış; farklı durum uzayı yapıları, faz planları, yaya bekleme süresi, kuyruk uzunluğu, gecikme ve CO_2 emisyonu birlikte değerlendirilmiştir. Böylece amaç, gerçek bir kavşak üzerinde araç akışı, yaya hareketi ve çevresel etkiler açısından daha dengeli bir trafik ışığı kontrol yaklaşımı geliştirmektir.

II. TRAFİK KAVŞAK MODELİ

A. *Ata Bulvarı-Fethiye Kavşağı*

Kavşak modeli oluşturulurken ilk olarak Ata Bulvarı-Fethiye Kavşağı'nın konum bilgisi SUMO'nun OpenStreetMap desteği kullanılarak alınmıştır. Harita verisi doğrudan kullanılmamış; çalışma dışı kalan yan yollar temizlenmiş, ana kavşak kolları korunmuş ve trafik ışığı kontrolünün uygulanacağı alan sadeleştirilmiştir. Bu düzenleme, simülasyonun hem daha anlaşılır hem de eğitim sürecinde daha kararlı çalışmasını sağlamıştır. Şekil 1'de görüldüğü gibi simülasyon ortamına kavşağın harita verisi çekilmiştir.



Şekil 1. SUMO ortamında modellenen kavşak yapısının genel görünümü.

B. *SUMO ve TraCI Tabanlı Benzetim Ortamı*

SUMO ve TraCI bağlantısı sayesinde simülasyon her adımda okunabilir ve kontrol edilebilir hâle getirilmiştir. Bu yapı ile şeritlerdeki araç sayısı, bekleme süresi, kuyruk uzunluğu ve yaya talepleri alınmış; ajan seçtiği faz kararını yine TraCI üzerinden simülasyona aktarmıştır. Böylece trafik ortamı ile

öğrenme algoritması arasında gerçek zamanlı bir veri alışverişi kurulmuştur.

III. PEKİŞTİRMELİ ÖĞRENME TABANLI TRAFİK IŞIĞI KONTROL YÖNTEMİ

Bu çalışmada pekiştirmeli öğrenme yapısı; ajan, ortam, durum, aksiyon ve ödül bileşenleri üzerinden kurulmuştur. Ajan, trafik ışığı kararını veren DQN tabanlı kontrol birimini ifade etmektedir. Ortam, SUMO üzerinde oluşturulan Ata Bulvarı-Fethiye Kavşağı modelidir. Durum, ajanın karar vermeden önce kavşaktan aldığı sayısal gözlem bilgisini; aksiyon ise seçilen trafik ışığı fazını ifade etmektedir.

Ajan her karar adımında mevcut trafik durumunu okumakta, uygun faz planı içerisinden bir aksiyon seçmekte ve bu seçimin sonucunda oluşan bekleme süresi, kuyruk uzunluğu, yaya bekleme süresi ve emisyon değerlerine göre ödül veya ceza almaktadır. Bu yapı sayesinde trafik ışığı kontrolü sabit zamanlı bir döngü yerine, anlık trafik koşullarına göre öğrenen bir karar verme problemine dönüştürülmüştür.

A. Durum Uzayı Senaryoları

Bu çalışmada durum, ajanın karar vermeden önce kavşak hakkında aldığı sayısal gözlem bilgisini ifade etmektedir. Başka bir deyişle durum; araç varlığı, şerit yoğunluğu, kapasite oranı ve trafik akışının modele giriş olarak aktarılmış hâlidir. Bu nedenle durum uzayı, pekiştirmeli öğrenme ajanının kavşağı hangi ayrıntı düzeyinde gördüğünü belirleyen temel veri yapısıdır. Bu kapsamda çalışmada üç farklı durum temsili karşılaştırılmıştır.

TABLO I. DURUM UZAYI SENARYOLARININ KARŞILAŞTIRILMASI

Durum	Kullanılan bilgi	Açıklama
Durum 1	80 hücrelik araç varlığı	Dört ana kol hücrelere ayrılmıştır. Araç bulunan hücre 1, boş hücre 0 olarak kodlanmıştır. Ayrıntılı temsil sağlar ancak giriş boyutu büyüktür
Durum 2	8 ana şerit grubu için toplam araç sayısı	Durum 1'deki hücre bilgisi sıkıştırılmıştır. Trafik yoğunluğu daha özet ve daha takip edilebilir biçimde modele aktarılır.
Durum 3	Şerit bazlı araç sayımı ve kapasite oranı	Sanal hücreleme yerine şerit doluluk oranı kullanılır. Daha sade, fiziksel olarak anlamlı ve düşük gürültülü bir gözlem yapısı sunar.

Genel olarak Tablo I incelendiğinde Durum 1 kavşağı ayrıntılı biçimde temsil ederken, Durum 2 daha özet bir trafik yoğunluğu bilgisini ifade etmektedir. Durum 3 ise daha sade, düşük gürültülü ve fiziksel olarak anlamlı bir doluluk bilgisi sunmaktadır.

B. Faz Planı Senaryoları

Bu çalışmada senaryo, Tablo II'ye bakıldığında trafik ışıklarının hangi sırayla ve hangi kontrol mantığıyla çalıştırılacağını belirleyen faz planını ifade etmektedir. Yani durum, ajanın kavşağı nasıl gördüğünü; faz planı ise ajanın hangi ışık düzenleri arasından seçim yaptığını göstermektedir. Bu nedenle durum uzayı ve faz planı birbirinden ayrı değerlendirilmiş, model performansı her iki yapı üzerinden karşılaştırılmıştır.

TABLO II. FAZ PLANI SENARYOLARININ KARŞILAŞTIRILMASI

Faz planı	Kontrol mantığı	Açıklama
Faz Planı 1	Simetrik ana yön kontrolü	Kuzey-güney ve doğu-batı doğrultularına dengeli geçiş hakkı verir. Daha sade yapısı nedeniyle eğitimde kararlı sonuçlar üretmiştir.
Faz Planı 2	Sıralı yön kontrolü	Tüm yönlerin teker teker geçiş aldığı daha ayrıntılı bir faz sıralaması kullanır. Trafiği ayırır ancak bekleme süresi bazı durumlarda artabilir
Faz Planı 3	Araç-yaya odaklı karma kontrol	Araç akışı ile yaya geçişlerini birlikte değerlendiren hibrit bir yapı sunar. Özellikle yaya bekleme süresi açısından daha dengeli sonuçlar verebilir

C. Ödül Fonksiyonu

Pekiştirmeli öğrenmede ödül fonksiyonu, ajana seçtiği aksiyonun ne kadar başarılı olduğunu gösteren temel geri bildirim mekanizmasıdır. Bu çalışmada ödül fonksiyonu, ajanın seçtiği trafik ışığı kararının kavşak üzerindeki etkisini sayısal olarak değerlendirmek amacıyla oluşturulmuştur. Ödül yapısının temelinde, bir önceki karar adımındaki toplam bekleme süresi ile yeni karar sonrasında oluşan toplam bekleme süresi arasındaki fark yer almaktadır. Bu ilişki denklem (1)'de gösterilmektedir. Denklemde (1) gösterilen R_t anlık ödül veya ceza değerini, W_{t-1} ise bir önceki adımdaki toplam bekleme süresini, W_t ise ışık kararı uygulandıktan sonraki yeni toplam bekleme süresini göstermektedir.

$$R_t = W_{t-1} - W_t \tag{1}$$

Denklem (1)'de gösterilen tanıma göre ışık kararı sonrasında toplam bekleme süresi azalırsa pozitif ödül oluşmakta, bekleme süresi artarsa ajan ceza almaktadır. Böylece model yalnızca belirli bir yöndeki araçları boşaltmaya değil, kavşağın genel bekleme yükünü azaltmaya yönlendirilmektedir. Yaya bekleme süresi, kuyruk uzunluğu ve emisyon etkisi de değerlendirme sürecinde dikkate alınarak sistemin sadece araç odaklı değil, genel trafik performansını iyileştiren bir kontrol davranışı öğrenmesi hedeflenmiştir.

D. DQN Yaklaşımı

Bu çalışmada trafik ışığı kontrolü için derin Q-öğrenme yaklaşımı kullanılmıştır. DQN, klasik Q-öğrenme yönteminin yapay sinir ağı ile desteklenmiş hâlidir. Klasik Q-öğrenmede her durum ve aksiyon çifti için Q değerleri tablo üzerinde tutulurken, trafik kavşağı gibi çok sayıda farklı durumun oluştuğu problemlerde bu yapı yetersiz kalabilmektedir. Bu nedenle çalışmada Q değerlerini tabloyla saklamak yerine, sinir ağı yardımıyla tahmin eden DQN modeli tercih edilmiştir.

DQN yapısında ajan, kavşaktan aldığı araç yoğunluğu, kuyruk durumu, şerit doluluk oranı ve durum uzayı bilgilerini sinir ağına giriş olarak aktarmaktadır. Sinir ağı bu bilgileri işleyerek her faz planı için beklenen Q değerlerini üretmektedir. Ajan ise bu değerlere göre o anki trafik durumu için en uygun faz kararını seçmeye çalışmaktadır.

Eğitim sürecinde ajan başlangıçta farklı faz kararlarını deneyerek çevreden ödül veya ceza almaktadır. Kuyruk

uzunluğu, bekleme süresi ve gecikme arttığında olumsuz geri bildirim oluşurken; trafik akışının düzenlenmesi ve bekleme süresinin azalması olumlu sonuç olarak değerlendirilmiştir. Böylece ajan, zamanla hangi trafik durumunda hangi faz planının daha uygun olduğunu öğrenmektedir. Bu yapı sayesinde DQN modeli, sabit zamanlı sistemlerden farklı olarak her yöne sabit süre vermek yerine kavşaktaki anlık yoğunluğa göre karar verebilmektedir. Böylece değişken trafik koşullarında daha esnek ve dengeli bir trafik ışığı kontrolü sağlanması amaçlanmıştır.

E. *Model Mimarisi ve Eğitim Parametreleri*

Eğitim sürecinde her bölüm bir epizot olarak değerlendirilmiştir. Her epizotta araç gelişleri yeniden oluşturulmuş ve ajan belirli zaman adımlarında kavşak durumunu okuyarak faz seçmiştir. Model mimarisi ve eğitim parametreleri, öğrenmenin kararlı ilerlemesi ve aşırı uyumun azaltılması amacıyla sade tutulmuştur. Bu kapsamda Tablo III'te görüldüğü gibi dört katmanlı derin sinir ağı yapısı, 400 nöron genişliği, 0.0001 öğrenme hızı, 0.75 gamma değeri ve 300 epizotluk eğitim süreci kullanılmıştır.

TABLO III. DQN MODEL MİMARİSİ VE EĞİTİM PARAMETRELERİ

Mimari	4 Katmanlı derin sinir ağı ve 400 nöron genişliği kullanılmıştır.
Gamma	0.75
Eğitim Bölümü	Model 300 simülasyon bölümü öğretilmiştir.
Öğrenme Hızı	0.0001

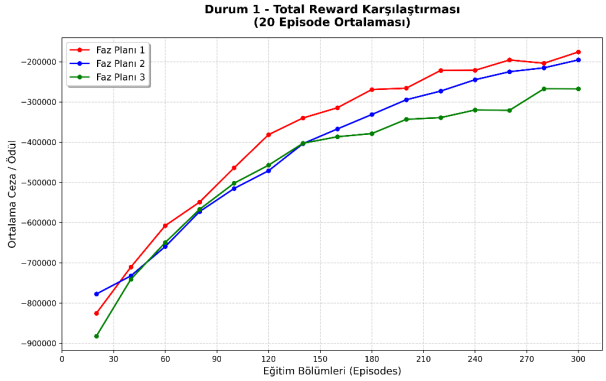
Bu kurgu sayesinde farklı durum uzayı ve faz planlarının aynı metrikler üzerinden karşılaştırılması mümkün olmuştur. Karşılaştırmada toplam ödül, CO_2 emisyonu, gecikme ve kuyruk uzunluğu değerleri kullanılmıştır. Böylece hangi durum temsilinin ajana daha anlaşılır bilgi verdiği ve hangi senaryo yapısının kavşak performansını daha dengeli iyileştirdiği incelenmiştir. Bu dört ölçüt birlikte değerlendirildiğinde modelin yalnızca matematiksel olarak değil, trafik yönetimi açısından da anlamlı olup olmadığı tartışılmıştır.

IV. BENZETİM SONUÇLARI VE KARŞILAŞTIRMALAR

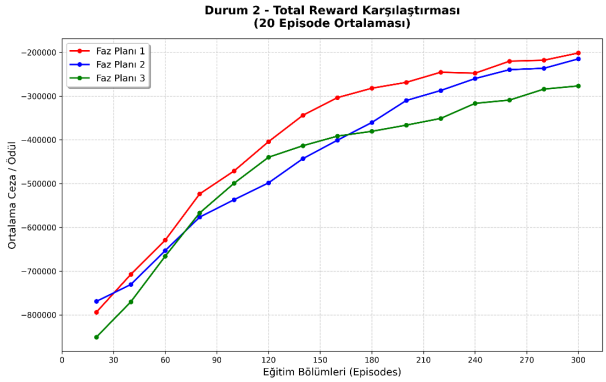
Bu kısımda, geliştirilen çalışmanın modelinin farklı durum uzayı ve faz planı senaryoları altında elde ettiği benzetim sonuçları değerlendirilmiştir. Karşılaştırmalarda toplam ödül, yaya bekleme süresi, kuyruk uzunluğu, gecikme, CO_2 emisyonu ve ortalama hız gibi performans ölçütleri dikkate alınmıştır. Bunun sonucunda yalnızca modelin öğrenme başarısı değil, aynı zamanda trafik akışı, yaya geçişleri ve çevresel etkiler üzerindeki etkisi de incelenmiştir. İlk olarak eğitim sürecinde faz planlarının toplam ödül değerleri karşılaştırılmış, ardından yaya bekleme süresi ve test senaryosunda durum uzayı performansları değerlendirilmiştir. Son aşamada ise en başarılı DQN tabanlı kontrol yapısı sabit zamanlı trafik ışığı kontrol sistemiyle karşılaştırılarak önerilen yöntemin kavşak yönetimindeki etkisi ortaya konulmuştur.

A. *Eğitim Sürecinde Faz Planı Karşılaştırması*

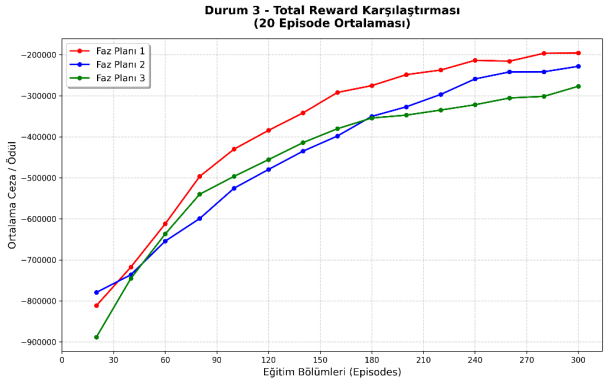
Benzetim sonuçlarının ilk aşamasında üç farklı durum uzayı için faz planlarının eğitim performansı karşılaştırılmıştır. Toplam ödül grafikleri, ajanın eğitim süreci boyunca kavşak yönetiminde ne kadar başarılı karar verdiğini göstermektedir. Ödül değerinin daha yüksek seviyelere yaklaşması, bekleme süresi, kuyruk oluşumu ve yaya cezası gibi olumsuz etkilerin azaldığını ifade etmektedir. Durum 1, Durum 2 ve Durum 3 için Faz planlarına göre ortalama ceza karşılaştırmalı sonuçları sırası ile şekil 1, şekil 2 ve şekil 3 eğrilerinde görülmektedir.



Şekil 2. Durum 1 için faz planlarına göre toplam ödül karşılaştırması.



Şekil 3. Durum 2 için faz planlarına göre toplam ödül karşılaştırması.



Şekil 4. Durum 3 için faz planlarına göre toplam ödül karşılaştırması.

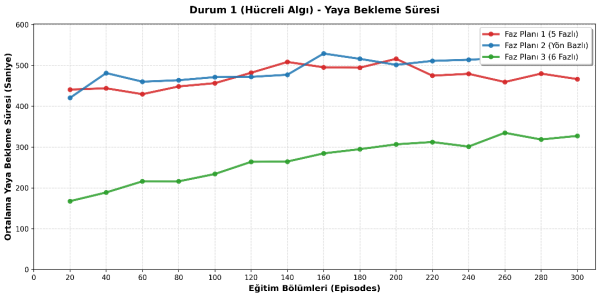
Durum 1, Durum 2 ve Durum 3 için elde edilen toplam ödül grafikleri birlikte değerlendirildiğinde eğitim sonunda en iyi faz planının genel olarak Faz Planı 1 olduğu görülmektedir. Faz Planı 1, daha düzenli ve sade bir geçiş mantığı sunduğu için ajanın öğrenme sürecinde daha kararlı sonuçlar üretmesini sağlamıştır. Faz Planı 2 ve Faz Planı 3 bazı bölümlerde iyileşme gösterse de özellikle karmaşık faz geçişleri nedeniyle ödül değerlerinde daha dalgalı bir yapı oluşmuştur.

Durum yapıları açısından bakıldığında Durum 1 ayrıntılı hücre bilgisi içerdiği için kavşaktaki araç konumlarını daha detaylı temsil etmektedir. Ancak bu ayrıntı giriş boyutunu artırdığı için öğrenme sürecinde dalgalanma oluşabilmektedir. Durum 2, hücre bilgisini şerit grupları düzeyinde

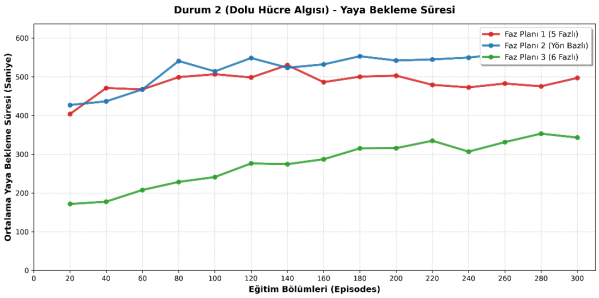
özetlediğinden daha sade bir öğrenme ortamı sunmuştur. Durum 3 ise doğrudan şerit bazlı araç sayımı ve kapasite oranını kullandığı için fiziksel olarak daha anlamlı bir temsil sağlamıştır.

B. Yaya Bekleme Süresi Sonuçları

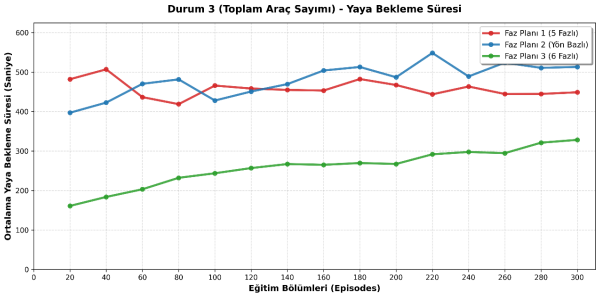
Yaya bekleme süresi grafikleri, faz planlarının yalnızca araç akışı açısından değil, yaya geçişleri açısından da değerlendirilmesi gerektiğini göstermektedir. Çalışmada bazı faz planlarında araç akışı iyileşirken yaya bekleme süresinin arttığı görülmüştür. Bunun nedeni, ajanın ödül yapısında araç kuyruğu ve araç gecikmesi gibi değerleri azaltmaya çalışırken bazı durumlarda yaya geçişlerini daha geç önceliklendirmesidir. Durum 1, Durum 2 ve Durum 3 için Faz planlarına göre yaya bekleme süreleri sırası ile şekil 1, şekil 2 ve şekil 3 eğrilerinde gösterilmektedir.



Şekil 5. Durum 1 için yaya bekleme süresi karşılaştırması.



Şekil 6. Durum 2 için yaya bekleme süresi karşılaştırması.



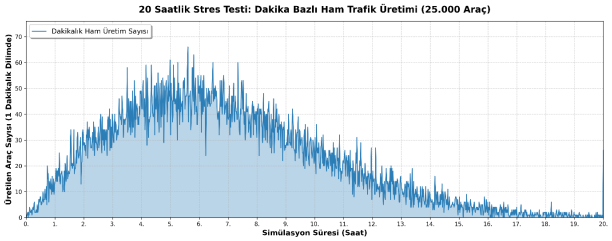
Şekil 7. Durum 3 için yaya bekleme süresi karşılaştırması.

Yaya bekleme süresi sonuçları incelendiğinde, özellikle Faz Planı 3'ün yaya geçişlerini daha dengeli ele aldığı görülmektedir. Bu faz planında araç akışı ile yaya geçişleri birlikte değerlendirildiği için yayaların kavşakta uzun süre beklemesi belirli ölçüde azaltılmıştır. Bu nedenle yaya bekleme süresi açısından en avantajlı faz planı Faz Planı 3 olarak yorumlanmıştır.

Bu sonuçlara göre yalnızca araç performansına bakarak en iyi faz planını seçmek yeterli değildir. Yaya bekleme süreleri de dikkate alındığında, faz planının sade olması kadar yaya güvenliğini ve geçiş önceliğini de koruması gerektiği anlaşılmaktadır.

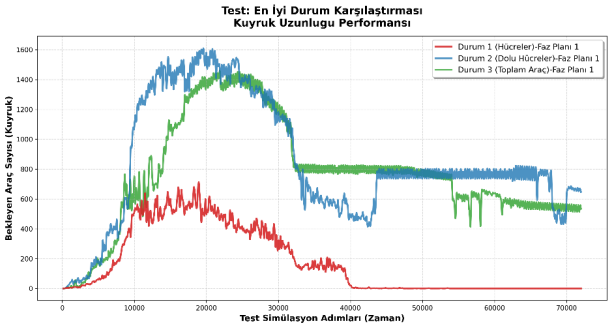
C. Test Senaryosunda Durum Uzayı Karşılaştırması

Eğitim sürecinden sonra model, Şekil 8'de görüldüğü gibi 20 saatlik ayrı bir test senaryosu üzerinde değerlendirilmiştir. Test sürecinde kullanılan trafik dağılımı, gün içerisindeki araç yoğunluğunun zamanla değişmesini temsil edecek şekilde oluşturulmuştur. Böylece model yalnızca eğitim ortamında değil, farklı yoğunluk seviyeleri içeren daha uzun süreli bir benzetim üzerinde de incelenmiştir.



Şekil 8. Test sürecinde kullanılan 20 saatlik trafik üretim dağılımı.

Eğitim sonuçlarında en iyi faz planı olarak Faz Planı 1 belirlendiği için test aşamasında durum yapıları bu faz planı altında karşılaştırılmıştır. Bu karşılaştırmanın amacı, aynı faz planı kullanıldığında ajanın kavşağı hangi durum temsiliyle daha başarılı yönettiğini belirlemektir.

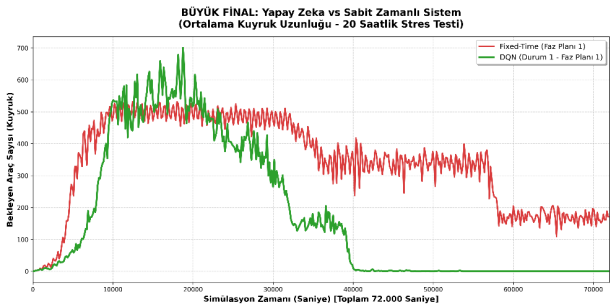


Şekil 9. En iyi faz planı altında durum yapılarının kuyruk uzunluğu performansı.

Şekil 9'da kuyruk uzunluğu karşılaştırması, Durum 1'in test sürecinde daha başarılı sonuç verdiğini göstermektedir. Durum 1'de araçların kavşağa olan konumu hücresel olarak temsil edildiği için ajan yalnızca toplam araç sayısını değil, araçların ışığa yakınlığını da değerlendirebilmiştir. Bu durum, özellikle yoğunluğun arttığı zaman aralıklarında daha doğru faz seçimi yapılmasına yardımcı olmuştur. Durum 2 ve Durum 3 ise daha sade yapılar olduğu için öğrenme sürecini kolaylaştırmış; ancak testte bazı yoğunluk değişimlerini Durum 1 kadar ayrıntılı yakalayamamıştır

D. Sabit Zamanlı Sistem ile DQN Karşılaştırması

Son aşamada en iyi sonuç veren yapı, sabit zamanlı trafik ışığı kontrolü ile karşılaştırılmıştır. Bu karşılaştırmada Şekil 10'da görüldüğü gibi Durum 1 ve Faz Planı 1 birleşimi pekiştirmeli öğrenme tabanlı kontrol yapısı olarak alınmıştır. Sabit zamanlı sistemde ışık süreleri trafik yoğunluğuna göre değişmediği için yoğun saatlerde kuyruk ve gecikme değerleri artabilmektedir. Pekiştirmeli öğrenme tabanlı sistem ise kavşaktan aldığı anlık durum bilgisine göre faz seçimi yaptığı için değişken trafik koşullarına daha iyi uyum sağlayabilmektedir.



Şekil 10. Pekiştirmeli öğrenme kontrolü ile sabit zamanlı sistemin ortalama kuyruk karşılaştırması.

Sabit zamanlı sistem ile pekiştirmeli öğrenme tabanlı kontrolün daha net karşılaştırılabilmesi için test sonuçları tablo hâlinde verilmiştir. Bu karşılaştırmada ortalama kuyruk uzunluğu, ortalama gecikme, toplam CO_2 emisyonu ve ortalama hız değerleri birlikte değerlendirilmiştir.

TABLO IV. DQN TABANLI KONTROL YAPISINA AİT TEST SONUÇLARI

Strateji (Durum-Faz Planı)	Ort. Kuyruk (Araç)	Ort.Gecikme (sn)	Toplam CO_2 (kg)	Ort.Hız (m/s)
Durum 1-Faz Planı 1	175.35	260895.24	31098.13	2.64
Durum 1-Faz Planı 2	181.12	143064.79	31905.63	3.31
Durum 1-Faz Planı 3	351.58	1722967.10	52419.76	1.29
Durum 2-Faz Planı 1	843.80	4306643.35	94087.91	0.83
Durum 2-Faz Planı 2	217.31	207610.58	37540.66	1.98
Durum 2-Faz Planı 3	860.93	6047077.40	91590.76	0.69
Durum 3-Faz Planı 1	774.56	6862544.51	84575.76	0.81
Durum 3-Faz Planı 2	333.79	523430.84	39922.87	2.71
Durum 3-Faz Planı 3	882.63	9054196.83	89838.39	0.41

Tablo IV incelendiğinde DQN tabanlı kontrol yapısında en dengeli sonucun Durum 1 - Faz Planı 1 ile elde edildiği görülmektedir. Bu yapıda ortalama kuyruk uzunluğu 175.35 araç, toplam CO_2 değeri ise 31098.13 kg olarak gerçekleşmiştir. Gecikme ve hız açısından Durum 1 - Faz Planı 2 daha avantajlı görünse de kuyruk uzunluğu ve emisyon değerleri birlikte değerlendirildiğinde Durum 1 - Faz Planı 1 daha dengeli bir kontrol davranışı ortaya koymuştur.

TABLO V. SABİT ZAMANLI KONTROL YAPISINA AİT TEST SONUÇLARI

Strateji (Fixed-Time)	Ort. Kuyruk (Araç)	Ort. Gecikme (sn)	Toplam CO_2 (kg)	Ort. Hız (m/s)
Faz Planı1	342.46	288340.48	50937.09	1.46
Faz Planı2	346.12	293058.48	51099.86	1.45
Faz Planı3	402.88	396565.70	56755.33	1.25

Tablo V'de verilen sabit zamanlı kontrol sonuçlarıyla karşılaştırıldığında, DQN tabanlı yapının özellikle kuyruk uzunluğu ve CO_2 emisyonu bakımından daha başarılı olduğu anlaşılmaktadır. Örneğin sabit zamanlı Faz Planı 1'de ortalama kuyruk 342.46 araç ve toplam CO_2 50937.09 kg iken, DQN'de Durum 1 - Faz Planı 1 için bu değerler sırasıyla 175.35 araç ve 31098.13 kg seviyesine düşmüştür. Bu sonuç, pekiştirmeli öğrenme tabanlı sistemin trafik durumuna göre daha esnek karar verebildiğini ve sabit süreli sisteme göre daha verimli bir akış sağladığını göstermektedir.

Tablolardan elde edilen sonuçlar, pekiştirmeli öğrenme tabanlı kontrolün sabit zamanlı sisteme göre daha dengeli bir trafik akışı sağladığını göstermektedir. Özellikle yoğunluğun arttığı durumlarda sabit zamanlı sistem her faza önceden belirlenmiş süre verdiği için kavşaktaki değişken trafik durumuna aynı hızda uyum sağlayamamaktadır.

CO_2 değerleri açısından bakıldığında DQN tabanlı kontrolün daha düşük emisyon oluşturduğu görülmektedir. Bu durum, araçların daha az beklemesi ve dur-kalk davranışının azalmasıyla ilişkilidir. Dolayısıyla önerilen kontrol yapısı yalnızca trafik performansı açısından değil, çevresel etki bakımından da sabit zamanlı yaklaşıma göre daha olumlu sonuç vermiştir.

Genel olarak benzetim sonuçları, Ata Bulvarı-Fethiye Kavşağı için pekiştirmeli öğrenme tabanlı trafik ışığı kontrolünün sabit zamanlı sisteme göre daha uyarlanabilir bir yapı sunduğunu göstermektedir. Eğitim aşamasında Faz Planı 1'in daha kararlı öğrenme sağladığı, test aşamasında ise Durum 1'in ayrıntılı hücresel temsil sayesinde daha başarılı kuyruk performansı verdiği görülmüştür. Bu nedenle çalışmada en başarılı yapı Durum 1 ve Faz Planı 1 birleşimi olarak değerlendirilmiştir.

V. SONUÇLAR VE GELECEK ÇALIŞMALAR

Bu çalışmada Ata Bulvarı-Fethiye Kavşağı için pekiştirmeli öğrenme tabanlı trafik ışığı kontrolü SUMO ortamında incelenmiştir. Çalışmanın temel amacı, sabit süreli sinyalizasyon yerine kavşaktaki anlık trafik durumuna göre karar verebilen DQN tabanlı bir kontrol yapısı geliştirmek ve bu yapıyı farklı durum uzayı ile faz planı senaryoları üzerinden değerlendirmektir. Bu kapsamda üç farklı durum temsili ve üç farklı faz planı karşılaştırılmış; toplam ödül, kuyruk uzunluğu, gecikme, yaya bekleme süresi ve CO_2 emisyonu gibi ölçütler üzerinden sistemin performansı incelenmiştir. Eğitim sonuçları genel olarak değerlendirildiğinde, Faz Planı 1'in daha kararlı bir öğrenme eğilimi gösterdiği görülmüştür. Bunun temel nedeni, Faz Planı 1'in daha sade ve düzenli bir geçiş mantığına sahip olmasıdır. Daha karmaşık faz yapılarında ajanın öğrenmesi zorlaşabildiği için ödül değerlerinde daha fazla dalgalanma oluşmuştur. Durum uzayı açısından bakıldığında ise Durum 1'in hücresel yapısı, araçların yalnızca sayısını değil kavşağa olan konumunu da modele aktardığı için test aşamasında daha başarılı sonuç vermiştir. Bu durum, özellikle yoğunluğun arttığı zaman aralıklarında ajanın daha doğru faz seçimi yapmasına katkı sağlamıştır.

Sabit zamanlı sistem ile DQN tabanlı kontrol karşılaştırıldığında, pekiştirmeli öğrenme yaklaşımının kuyruk uzunluğu ve CO_2 emisyonu bakımından daha başarılı olduğu görülmüştür. Sabit süreli sistemde ışık süreleri trafik yoğunluğuna göre değişmediği için yoğun saatlerde bazı kollarda gereksiz bekleme ve kuyruk oluşumu artmaktadır. DQN tabanlı sistem ise kavşaktan aldığı anlık durum bilgisine göre karar verdiğinden, trafik yoğunluğundaki değişimlere daha esnek tepki verebilmiştir. Bu durum, araçların daha az

beklemesine, dur-kalk davranışının azalmasına ve buna bağlı olarak emisyon değerlerinin düşmesine katkı sağlamıştır.

Yaya bekleme süresi sonuçları ise sistemin yalnızca araç akışı üzerinden değerlendirilmemesi gerektiğini göstermiştir. Faz Planı 3, bazı durumlarda yaya bekleme süresi açısından daha dengeli sonuçlar üretmiştir. Ancak bu faz yapısı daha karmaşık olduğu için genel araç performansı ve öğrenme kararlılığı açısından her durumda en iyi sonucu vermemiştir. Bu nedenle çalışmada en başarılı genel yapı Durum 1 ve Faz Planı 1 birleşimi olarak değerlendirilmiştir.

Genel olarak elde edilen sonuçlar, pekiştirmeli öğrenme tabanlı trafik ışığı kontrolünün sabit zamanlı sistemlere göre daha uyarlanabilir bir yapı sunduğunu göstermektedir. Bu yaklaşım, kavşaktaki anlık yoğunluk değişimlerini dikkate alarak araç akışını iyileştirme, kuyruk uzunluğunu azaltma ve çevresel etkileri düşürme potansiyeline sahiptir.

Gelecek çalışmalarda, geliştirilen DQN tabanlı trafik ışığı kontrol yapısının gerçek zamanlı veya gerçek veriye dayalı trafik akışlarıyla test edilmesi önerilmektedir. Böylece sistemin sabah-akşam yoğunluğu ve ani trafik artışları gibi değişken koşullardaki başarısı daha ayrıntılı incelenebilir. Ayrıca modelin komşu kavşakları da kapsayacak şekilde genişletilmesiyle çoklu kavşak koordinasyonu sağlanabilir. Ödül fonksiyonuna acil durum araçları, toplu taşıma önceliği ve farklı emisyon türleri gibi ölçütler eklenerek sistem daha kapsamlı hâle getirilebilir. Bunun yanında Double DQN, Dueling DQN ve çok ajanlı pekiştirmeli öğrenme gibi farklı algoritmalarla model performansı karşılaştırılabilir.

KAYNAKÇA

[1] X. Liang, X. Du, G. Wang, and Z. Han, “A deep reinforcement learning network for traffic light cycle control,” IEEE Transactions on Vehicular Technology, vol. 68, no. 2, pp. 1243–1253, 2019, doi: 10.1109/TVT.2018.2890726.

[2] H. Wei, H. Yao, G. Zheng, and Z. Li, “IntelliLight: A reinforcement learning approach for intelligent traffic light control,” in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, London, United Kingdom, 2018, pp. 2496–2505, doi: 10.1145/3219819.3220096.

[3] H. Wei et al., “PressLight: Learning max pressure control to coordinate traffic signals in arterial network,” in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Anchorage, AK, USA, 2019, pp. 1290–1298, doi: 10.1145/3292500.3330949.

[4] G. Zheng, Y. Xiong, J. Zang, J. Feng, J. Wei, and H. Wei, “Learning phase competition for network-wide traffic signal control,” in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Anchorage, AK, USA, 2019, pp. 1969–1979, doi: 10.1145/3292500.3330983.

[5] L. Wu, K. Wu, and J. Wang, “An empirical study of design elements in deep reinforcement learning for traffic signal control,” in Proceedings of the 2020 IEEE International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 2020, pp. 1–6, doi: 10.1109/ITSC45102.2020.9294245.

[6] X. Zhang, Y. Wang, and L. Li, “Fuzzy logic-based deep Q-learning for traffic signal control at intersections,” IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 4, pp. 2145–2156, Apr. 2021, doi: 10.1109/TITS.2020.2987654.

[7] C. Cai, Y. Wang, and G. Zhang, “A robust deep reinforcement learning model for traffic signal control with dueling double deep Q-networks and prioritized experience replay,” IEEE Access, vol. 9, pp. 110231–110242, 2021, doi: 10.1109/ACCESS.2021.3102456.

[8] Y. Wang, T. Xu, and T. Zhao, “EP-D3QN: An optimization method for traffic light timing based on deep reinforcement learning,” IEEE Access, vol. 9, pp. 132482–132493, 2021, doi: 10.1109/ACCESS.2021.3115456.

[9] J. Ault and G. Sharon, “Reinforcement learning benchmarks for traffic signal control,” in Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems, London, United Kingdom, 2021, pp. 1435–1437, doi: 10.5555/3463952.3464124.

[10] J. Zhang et al., “CityFlow: A multi-agent reinforcement learning environment for large scale city traffic scenario,” in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Anchorage, AK, USA, 2019, pp. 1969–1977, doi: 10.1145/3292500.3331025.

[11] Z. Li et al., “CoLight: Learning cooperative graph policy for network-wide traffic signal control,” in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Anchorage, AK, USA, 2019, pp. 2496–2505, doi: 10.1145/3292500.3330961.

[12] B. Khazaeni, M. R. Khazaeni, and A. A. Safaei, “Federated reinforcement learning for network-wide urban traffic signal control,” IET Intelligent Transport Systems, vol. 16, no. 8, pp. 1045–1058, Aug. 2022, doi: 10.1049/itr2.12195.

[13] H. Wei et al., “EcoLight: Reward shaping in deep reinforcement learning for environmentally-friendly traffic signal control,” in Proceedings of the 2021 IEEE International Conference on Smart Data (SmartData), 2021, pp. 312–319, doi: 10.1109/SmartData54321.2021.00042.

[14] M. S. Guler and T. S. Atan, “Pedestrian-friendly traffic signal control using SUMO and TraCI,” in Proceedings of the International Conference on Simulation and Modeling, Istanbul, Turkey, 2020, pp. 45–52.

[15] M. B. Aydoğan, E. Muslu, I. Tunc, and M. T. Söylemez, “Bursa şehri Şehreküstü ve Gökdere bölgeleri için derin Q öğrenme yöntemiyle trafik ışık kontrolü (Traffic light control with deep Q learning for Sehrekustu and Gokdere areas of Bursa city),” Conference Paper, 2024.

Yangın Anında Güvenli Tahliye İçin Pekiştirmeli Öğrenme Tabanlı Bir Yaklaşım

A Reinforcement Learning-Based Approach for Safe Evacuation During Fire Emergencies

Alptuğ YEL
BURSA TEKNİK ÜNİVERSİTESİ
Mekatronik Mühendisliği
Bursa, Türkiye
alptugyel.4309@gmail.com

Yağız CANER
BURSA TEKNİK ÜNİVERSİTESİ
Mekatronik Mühendisliği
Bursa, Türkiye
yskcaner12@gmail.com

Dr. Seyitali İLYAS
KIRGIZİSTAN-TÜRKİYE MANAS ÜNİVERSİTESİ
Elektrik ve Elektronik Mühendisliği Bölümü
Bishkek, Kırgızistan
seyitali.ilyas@manas.edu.kg

Dr. Öğr. Üyesi İlhan TUNÇ
BURSA TEKNİK ÜNİVERSİTESİ
Mekatronik Mühendisliği
Bursa, Türkiye
ilhan.tunc@btu.edu.tr

Özet — Kapalı ve çok katlı binalarda çıkan yangınlar, özellikle panik ortamında bireylerin doğru karar vermesini güçleştirerek tahliye süreçlerini tehlikeli bir hal almaktadır. Sabit kaçış tabelaları ve önceden belirlenmiş tahliye planları, dumanın hızla yayıldığı ya da bazı koridorların erişilemez hale geldiği dinamik yangın ortamlarında yetersiz kalmaktadır. Bu problemi ele almak amacıyla çalışmamızda, bina içindeki anlık sıcaklık ve duman dağılımına göre güvenli kaçış rotası hesaplayabilen pekiştirmeli öğrenme tabanlı bir karar destek ajanı geliştirilmiştir. Ajan, yangın bölgeleri ve duman yoğunluğunun insan sağlığına etkisini çevresel geri bildirim olarak alarak farklı senaryolarda güvenli tahliye politikalarını öğrenmektedir. Temel hedef, bireyin hem yangına hem de solunumu tehdit eden yoğun dumana en az maruz kalacağı rotayı bularak çıkış noktasına ulaşmasını sağlamaktır. Sunulan yöntem, gelecekte akıllı bina yönetim sistemlerine entegre edilebilecek, uyarlanabilir bir acil durum yönlendirme altyapısı olarak değerlendirilmektedir.

Anahtar Kelimeler: Pekiştirmeli Öğrenme, Güvenli Rota Optimizasyonu, Acil Durum Yönetimi

Abstract — Emergencies such as fires make evacuation processes critical, especially in enclosed and complex buildings. The inability of individuals to make sound decisions due to panic, stress, and environmental conditions during a fire complicates the evacuation process and increases the risk of casualties. In this study, a reinforcement learning-based decision support agent is proposed to ensure the safe and rapid evacuation of individuals during a fire. The developed method generates dynamically generated escape routes by considering the dynamically changing temperature and smoke propagation in the building environment. During the training process of the agent, fire zones, smoke density, and the effects of these elements on human health are considered as environmental parameters. The system is modeled to learn safe evacuation strategies by sensing the environmental state under various scenarios. In this context, an approach is presented that enables the individual to reach the safe zone as quickly as possible, taking into account the effects of fire and smoke they are exposed to. Thus, it is aimed for individuals to reach the safe exit point with minimal harm by being less affected by fire-induced negative factors such as high temperature and dense smoke. As a result, the developed method offers an intelligent and adaptable solution framework that can be used in emergency evacuation systems.

Keywords: Reinforcement Learning, Safe Route Optimization, Emergency Management.

I. GIRIŞ

Günümüz binalarında artan kat sayısı, birbirine geçen koridor ağları ve yüksek kullanıcı yoğunluğu, yangın anındaki tahliye yönetimini karmaşık bir mühendislik problemine dönüştürmektedir. Panik ve stres altındaki bireyler çoğunlukla alışılageldik kaçış yollarına yönelir; ancak dumanın veya alevlerin bu güzergahları kapattığı durumlarda söz konusu refleks can tehlikesi oluşturabilmektedir. Mevcut sistemlerde yaygın biçimde kullanılan sabit yönlendirme levhaları ve önceden belirlenmiş kaçış planları, yangının anlık gelişimine ayak uyduramamaktadır. Özellikle dumanın hızla dolduğu ya da yüksek sıcaklıkların belirli güzergahları kullanılamaz kıldığı senaryolarda bu statik çözümler bireyleri doğrudan tehlikeli bölgelere yönlendirebilmekte ve tahliye sürecini daha riskli hale getirebilmektedir.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Son yıllarda yapay zeka ve IoT tabanlı sensör teknolojilerindeki ilerleme, tahliye optimizasyonuna yönelik araştırmalara ivme kazandırmıştır. Literatürde bu tür problemler için sosyal kuvvet modelleri [1], hücresel otomatlar [2] ve karınca kolonisi optimizasyonu gibi sezgisel yöntemler yaygın olarak kullanılmaktadır. Ancak bu yöntemler, yangın ve duman yayılımı gibi zamana bağlı olarak sürekli değişen çevresel koşullarda gerçek zamanlı adaptasyon konusunda sınırlı kalmaktadır. Çevreden aldığı geri bildirimler doğrultusunda deneme-yanılma yöntemiyle öğrenebilen Pekiştirmeli Öğrenme algoritmaları ise dinamik, belirsiz ve doğrusal olmayan bu tür problemlerde daha esnek ve etkili çözümler sunmaktadır.

Bu çalışmada, karmaşık bina yapılarında meydana gelen yangın senaryoları için pekiştirmeli öğrenme tabanlı akıllı bir tahliye ajanı geliştirilmiştir. Tasarlanan sistemde ajan, duvarlar, ateş bölgeleri, duman alanları ve çıkış noktalarından oluşan grid tabanlı bir ortam içerisinde hareket etmektedir. Yangın ve dumanın zamanla yayılması sayesinde ortam dinamik hale getirilmiş ve gerçek hayata daha yakın bir simülasyon oluşturulmuştur. Ajanın temel amacı, tehlikeli bölgelerden mümkün olduğunca kaçınarak en güvenli rotayı öğrenmek ve en kısa sürede çıkış noktasına ulaşmaktır.

Önerilen modelde, çevreden alınan anlık geri bildirimler doğrultusunda çalışan ödül-ceza mekanizması kullanılmıştır. Ateş ve yoğun duman bölgelerine yaklaşılması durumunda negatif ödüller uygulanırken, güvenli hareketler ve çıkış noktasına ulaşılması pozitif ödüller ile desteklenmiştir. Böylece sistem yalnızca en kısa yolu değil, aynı zamanda insan sağlığını tehdit eden çevresel faktörleri de dikkate alan güvenli rotaları öğrenebilmektedir. Çalışmada Q-Learning ve SARSA algoritmaları kullanılarak ajan davranışları karşılaştırılmış, ayrıca keşif-sömürü dengesi için üstel epsilon azaltımı yöntemi uygulanmıştır.

Elde edilen sonuçlar, geliştirilen yapay zeka tabanlı sistemin yangın ve duman gibi dinamik tehlikeler karşısında güvenli tahliye rotalarını başarılı şekilde öğrenebildiğini göstermektedir. Bu yönüyle çalışma, gelecekte akıllı bina teknolojileri, afet yönetim sistemleri ve otonom acil durum yönlendirme uygulamaları için önemli bir altyapı sunmaktadır.

Bu çalışmanın temel katkıları üç başlık altında özetlenebilir. İlk olarak, yangın ve duman yayılımını zamana bağlı olarak güncelleyen dinamik grid tabanlı bir tahliye ortamı oluşturulmuştur. İkinci olarak, ajanın yalnızca en kısa yolu değil, aynı zamanda yangın ve duman kaynaklı riskleri dikkate alan güvenli rotaları öğrenebilmesi için güvenlik odaklı bir ödül-ceza mekanizması tasarlanmıştır. Üçüncü olarak, Q-Learning ve SARSA algoritmaları aynı simülasyon ortamında karşılaştırılarak güvenli tahliye başarımı açısından değerlendirilmiştir.

Çalışmanın devamı şu şekilde düzenlenmiştir. İkinci bölümde geliştirilen sistemin genel tasarımı, grid tabanlı ortam yapısı ve yangın-duman yayılım modeli açıklanmıştır. Üçüncü bölümde pekiştirmeli öğrenme tabanlı tahliye yaklaşımı, durum uzayı, aksiyon uzayı, ödül-ceza mekanizması ve kullanılan öğrenme algoritmaları sunulmuştur. Dördüncü bölümde benzetim çalışmaları ve algoritmaların performans karşılaştırmaları verilmiştir. Son bölümde ise elde edilen sonuçlar değerlendirilmiş ve gelecekte yapılabilecek çalışmalar tartışılmıştır.

II. SISTEM TASARIMI

Yangın tahliyesi, duman yayılımı, sıcaklık artışı ve görüş mesafesinin kısalması gibi zaman içinde sürekli değişen etkenler nedeniyle statik optimizasyon yaklaşımlarıyla çözülmesi güç, çok boyutlu bir problemdir. Geleneksel rota planlama yöntemleri bu dinamikleri yeterince modelleyememekte; ortam değiştikçe çözüm kalitesi hızla düşmektedir. Bu eksiklik, son yıllarda gerçek zamanlı karar verebilen öğrenme temelli sistemlere olan ilgiyi artırmıştır. Aşağıda çalışmanın dayandığı pekiştirmeli öğrenme çerçevesi, Q-Learning ve SARSA algoritmaları ile yangın tahliyesine ilişkin mevcut literatür özetlenmiştir.

Pekiştirmeli Öğrenme (Reinforcement Learning), bir ajanın çevresiyle doğrudan etkileşim kurarak deneme-yanılma yoluyla öğrenmesine dayanan bir makine öğrenmesi paradigmasıdır. Ajan, her eylem sonrasında aldığı ödül ya da ceza sinyallerine göre politikasını günceller ve zamanla toplam birikimli ödülü en yüksek kılacak davranışları öğrenir [3]. Ortam hakkında önceden bir model gerektirmeyen bu yapı, ajanın amacına giden en iyi karar dizisini keşfetmesine olanak tanımaktadır.

Temel bileşenler incelendiğinde; ajan karar veren birimi, ortam ise ajanın içinde bulunduğu simülasyon ya da gerçek dünya ortamını temsil etmektedir. Durum (s), belirli bir andaki çevresel koşulların anlık görüntüsüdür; aksiyon (a) ise ajanın o anda seçebildiği hareketleri kapsar. Ödül sinyali, gerçekleştirilen hareketin ne denli uygun olduğunu sayısal olarak ifade eder: güvenli bir adım pozitif, tehlikeli bir hareket negatif geri bildirim üretir. Zamanla bu geri bildirimler birikerek ajanın hangi eylemlerin ne ölçüde avantajlı olduğunu içselleştirmesini sağlar.

Bu yaklaşımın tahliye problemleri açısından kritik bir avantajı, önceden yazılmış kural setlerine gerek duymaksızın doğrudan çevreden öğrenebilmesidir. Yangın yayılırken bazı koridorlar kapanır, duman yoğunluğu dakikalar içinde değişir; bu tür dinamik ortamlarda önceden kodlanmış kurallar hızla geçerliliğini yitirir. Pekiştirmeli öğrenme ise her adımda gelen geri bildirimi değerlendirerek politikasını günceller, dolayısıyla sabit rota algoritmalarının yetersiz kaldığı koşullarda bile işlevini korur.

A. Q-Learning Algoritması

Q-Learning algoritması, model gerektirmeyen ve off-policy tabanlı bir pekiştirmeli öğrenme yöntemidir. Algoritma, ajanın gelecekte elde edebileceği maksimum ödülü dikkate alarak en uygun hareketleri öğrenmesini sağlamaktadır [4]. Bu yöntemde ajan, bulunduğu durum ve gerçekleştirdiği aksiyonlar doğrultusunda bir Q-değeri oluşturmakta ve zaman içerisinde bu değerleri güncelleyerek en iyi politikayı öğrenmektedir.

Q-Learning algoritmasının temel amacı, belirli bir durumda gerçekleştirilecek en uygun aksiyonu belirlemektir. Öğrenme süreci boyunca ajan, farklı hareketleri deneyerek çevreden geri bildirim almakta ve toplam ödülü maksimize edecek hareketleri tercih etmektedir. Algoritmanın güncelleme işlemi aşağıdaki eşitlik ile ifade edilmektedir:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)] \quad (1)$$

Denklem (1)'de α öğrenme katsayısını, γ indirim faktörünü ve r anlık ödülü temsil etmektedir. Q-Learning algoritması genellikle en yüksek ödülü sağlayan yolları

öğrenme eğilimindedir. Ancak bazı durumlarda kısa fakat riskli rotaları tercih edebilmesi, güvenlik odaklı uygulamalarda dikkat edilmesi gereken bir durum oluşturmaktadır.

B. SARSA Algoritması

SARSA algoritması, model gerektirmeyen ve on-policy tabanlı çalışan bir pekiştirmeli öğrenme yöntemidir. Q-Learning algoritmasından farklı olarak SARSA, güncelleme işlemini ajanın gerçekten seçtiği aksiyona göre gerçekleştirmektedir [5]. Bu nedenle algoritma, öğrenme süreci boyunca daha güvenli ve temkinli politikalar geliştirebilmektedir.

SARSA algoritmasında öğrenme süreci; mevcut durum, gerçekleştirilen aksiyon, alınan ödül, bir sonraki durum ve bir sonraki aksiyon bileşenlerinden oluşmaktadır. Algoritmanın güncelleme denklemi aşağıdaki şekilde ifade edilmektedir:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma Q(s',a') - Q(s,a)] \quad (2)$$

Denklem (2) sayesinde ajan, yalnızca teorik olarak en yüksek ödülü değil, uygulamada gerçekten takip ettiği politikayı öğrenmektedir. Literatürde yapılan çalışmalarda SARSA algoritmasının özellikle riskli bölgelerden kaçınma konusunda daha güvenli davranışlar sergilediği belirtilmiştir [5]. Bu nedenle yangın tahliye sistemleri gibi insan güvenliğinin ön planda olduğu uygulamalarda önemli avantajlar sunmaktadır.

C. Yangın Tahliye Sistemleri

Yangın tahliye sistemleri üzerine yapılan çalışmalarda, bireylerin acil durum anındaki hareket davranışlarını modellemek amacıyla farklı matematiksel ve simülasyon tabanlı yöntemler geliştirilmiştir. Bu alandaki önemli çalışmalardan biri, Dirk Helbing ve Péter Molnár yaya dinamiklerini sosyal kuvvet yaklaşımıyla modellemiştir [1]. Bu modelde yayaların hareketleri; hedefe ulaşma isteği, diğer bireylerden kaçınma davranışı ve çevresel engellerin etkisi gibi sosyal kuvvetler yardımıyla modellenmiştir. Çalışma, kalabalık insan davranışlarının simülasyonu ve acil durum tahliye analizleri açısından literatürde temel referanslardan biri olarak kabul edilmektedir.

Hücresel otomat tabanlı tahliye modelleri üzerine gerçekleştirilen önemli çalışmalardan biri ise Christian Burstedde, Kai Klauck, Andreas Schadschneider ve Johannes Zittartz iki boyutlu hücresel otomat kullanarak yaya dinamiklerinin simülasyonunu gerçekleştirmiştir [2]. Bu çalışmada yayaların hareketleri grid tabanlı bir yapı üzerinde modellenmiş ve özellikle dar koridorlar ile yoğun insan akışının bulunduğu ortamlardaki tahliye davranışları incelenmiştir. Geliştirilen model, düşük hesaplama maliyeti nedeniyle tahliye simülasyonlarında yaygın olarak kullanılmaktadır.

Tahliye modelleme yöntemlerinin genel yapısını inceleyen önemli çalışmalardan biri de Sharon Gwynne, Eric Galea, Michael Owen, Peter J. Lawrence ve Luís Filippidis yapılı çevrelerden tahliye için kullanılan bilgisayar simülasyon yöntemlerini değerlendirmiştir [6]. Bu çalışmada mevcut tahliye simülasyon yöntemleri karşılaştırılmış ve birçok sistemin statik çevre varsayımlarına dayandığı belirtilmiştir. Özellikle yangın sırasında ortaya çıkan duman yayılımı, sıcaklık değişimi ve kapanan kaçış yolları gibi dinamik çevresel faktörlerin klasik modeller tarafından yeterince temsil edilemediği vurgulanmıştır.

Son yıllarda ise yapay zekâ ve pekiştirmeli öğrenme tabanlı yöntemler, dinamik tahliye problemleri için güçlü bir alternatif olarak değerlendirilmektedir. Pekiştirmeli öğrenme algoritmaları, çevreden alınan anlık geri bildirimler doğrultusunda öğrenme gerçekleştirebildiği için yangın ve duman gibi sürekli değişen tehlikelerin bulunduğu ortamlarda daha adaptif çözümler sunabilmektedir. Bu nedenle yapay zekâ destekli tahliye sistemleri, akıllı bina teknolojileri ve afet yönetim uygulamaları açısından giderek daha önemli bir araştırma alanı haline gelmektedir.

III. Pekiştirmeli Öğrenme Tabanlı Yangın Tahliye Yöntemi

Bu çalışmada, yangın anında güvenli tahliye rotalarının oluşturulabilmesi amacıyla pekiştirmeli öğrenme tabanlı dinamik bir simülasyon ortamı geliştirilmiştir. Tasarlanan sistem, grid tabanlı bir yapı üzerinde çalışan ve çevresel değişimlere gerçek zamanlı olarak uyum sağlayabilen akıllı bir ajan yapısından oluşmaktadır. Simülasyon ortamında yangın, duman, duvar ve çıkış noktaları gibi farklı çevresel bileşenler tanımlanmış; ajanın bu değişken koşullar altında güvenli tahliye stratejileri öğrenmesi hedeflenmiştir.

Geliştirilen sistemde ajan, çevreden aldığı geri bildirimler doğrultusunda hareket ederek ödül-ceza mekanizması yardımıyla öğrenme gerçekleştirmektedir. Yangın ve duman yayılımının zamanla değişmesi sayesinde ortam statik olmaktan çıkarılmış ve gerçek hayattaki acil durum senaryolarına daha yakın bir yapı oluşturulmuştur. Böylece ajan yalnızca en kısa yolu değil, aynı zamanda insan güvenliğini ön planda tutan optimum kaçış rotalarını öğrenebilmektedir.

Bu bölümde geliştirilen simülasyon ortamının yapısı, yangın ve duman yayılım mekanizmaları, ajan hareket sistemi, ödül-ceza yapısı ve keşif-sömürü stratejileri detaylı olarak açıklanmıştır.

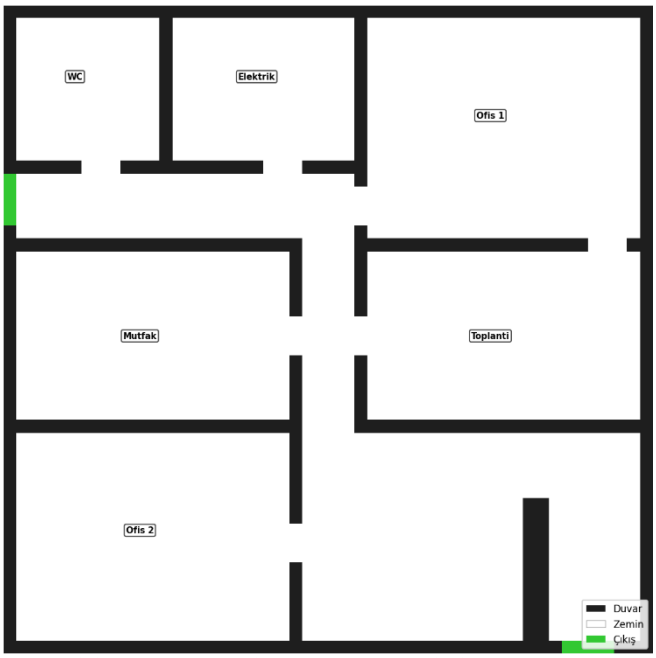
A. Ortam Tasarımı

Geliştirilen simülasyon ortamı, 50x50 boyutunda iki boyutlu grid tabanlı bir yapı kullanılarak oluşturulmuştur. Ortam içerisindeki her hücre belirli bir çevresel durumu temsil etmektedir. Bu yapı içerisinde duvarlar, boş alanlar, yangın bölgeleri, duman alanları, çıkış noktaları ve ajan konumu ayrı hücre tipleri olarak tanımlanmıştır. Grid tabanlı yapı sayesinde ajan, çevreyi hücre bazlı olarak algılayabilmekte ve karar verme süreçlerini bulunduğu duruma göre gerçekleştirebilmektedir.

Simülasyon ortamında duvar hücreleri geçilemez bölgeleri temsil ederken, çıkış hücreleri ajanın ulaşması gereken hedef noktaları ifade etmektedir. Yangın hücreleri yüksek riskli bölgeleri temsil etmekte ve ajan için ciddi negatif ödüller oluşturmaktadır. Duman hücreleri ise görüş kaybı ve insan sağlığı üzerindeki olumsuz etkiler nedeniyle orta seviyede riskli bölgeler olarak değerlendirilmiştir. Boş alanlar ise ajanın serbest şekilde hareket edebildiği güvenli bölgeleri temsil etmektedir.

Ajan, ortam içerisinde yukarı, aşağı, sağ ve sol olmak üzere dört farklı yönde hareket edebilmektedir. Her hareket sonrasında ortam yeniden değerlendirilmekte ve yangın ile duman yayılımı güncellenmektedir. Böylece ajan, sürekli

değişen çevre koşulları altında karar vermek zorunda kalmaktadır. Bu dinamik yapı sayesinde sistem, gerçek yangın senaryolarına daha yakın bir simülasyon ortamı sunmaktadır.



Şekil.1. Simülasyonda kullanılan 50×50 grid tabanlı ofis ortamı

B. Yangın ve Duman Yayılım Modeli

Geliştirilen sistemde ortamın dinamik hale getirilebilmesi amacıyla yangın ve duman yayılım mekanizmaları modele dahil edilmiştir. Yangın ve duman yayılım katsayıları, sistemin kararlı çalışması ve gerçekçi dinamik davranışlar üretebilmesi amacıyla deneysel olarak belirlenmiştir. Parametre seçim sürecinde farklı yayılım olasılıkları test edilmiş; çok düşük değerlerde yangının ortam üzerinde yeterli etki oluşturmadığı, çok yüksek değerlerde ise ortamın kısa sürede tamamen erişilemez hale geldiği gözlemlenmiştir. Bu durumların sonucunda yangın yayılımı, stokastik ve komşuluk tabanlı bir yapı kullanılarak modellenmiştir. Simülasyonda başlangıç yangınlarının çevre hücrelere yayılma olasılığı %35 ($p_{fire_spread} = 0.35$) olarak belirlenmiştir. Bununla birlikte sistem yalnızca tek bir yayılım olasılığına bağlı çalışmamaktadır. Yangına maruz kalan her hücrenin çevresindeki komşu hücreler ayrıca değerlendirilmekte ve her aktif yangın hücresi komşu hücreler üzerinde ek yayılım etkisi oluşturmaktadır. Bu komşu etkileşim katsayısı $p_{neighbor_burn} = 0.10$ olarak tanımlanmıştır.

$$P_{total} = \min\left(0.35\ x\ \left(N_{fire} \times 0.10\right)\ , 1.0\right) \qquad (3)$$

Denklem (3)'de P_{total}toplam yangın yayılım olasılığını, N_{fire} ise ilgili hücrenin çevresindeki aktif yangın komşularının sayısını temsil etmektedir. Bu yapı sayesinde çevresinde daha fazla aktif yangın hücresi bulunan bölgelerde yayılım riski artmakta ve yangının davranışı daha gerçekçi şekilde modellenmektedir.

Bu yapı sayesinde bir hücrenin yanma ihtimali yalnızca başlangıç yayılım olasılığına değil, çevresindeki aktif yangın hücrelerinin yoğunluğuna da bağlı hale getirilmiştir. Özellikle birden fazla yanan hücre ile çevrili bölgelerde yangının yayılma riski artmakta ve simülasyon ortamı daha gerçekçi bir davranış sergilemektedir. Her zaman adımında aktif yangın hücrelerinin çevresindeki uygun hücreler yeniden değerlendirilmekte ve yangın yayılım haritası güncellenmektedir. Böylece ortam sürekli değişmekte ve ajanın sabit bir rota ezberlemesi engellenmektedir.

Duman yayılım modeli ise yangına bağlı olarak genişleyen bir risk alanı oluşturacak şekilde tasarlanmıştır. Duman yayılım hızı $smoke_speed = 0.6$ katsayısı kullanılarak modellenmiş ve her zaman adımında duman yarıçapı dinamik olarak artırılmıştır. Böylece yangın bölgelerinden yayılan duman, belirli bir yarıçap içerisinde çevre hücrelere yayılmaktadır. Ajan yalnızca doğrudan ateşten değil, aynı zamanda görüş kaybı ve solunum riski oluşturan duman bölgelerinden de kaçınmak zorunda kalmaktadır. Duman yayılımının sisteme dahil edilmesi, gerçek yangın senaryolarındaki çevresel risklerin daha doğru modellenmesine katkı sağlamaktadır.

Yangın ve dumanın zamanla değişen yapısı sayesinde ortam statik olmaktan çıkarılmış ve gerçek zamanlı karar verme gerektiren dinamik bir simülasyon elde edilmiştir. Bu yapı, pekiştirmeli öğrenme algoritmalarının farklı senaryolara uyum sağlayabilme yeteneğini değerlendirmek açısından önemli avantaj sağlamaktadır.

C. Ajan Yapısı ve Hareket Mekanizması

Geliştirilen sistemde ajan, grid tabanlı ortam içerisinde hareket ederek güvenli çıkış rotasını öğrenmeye çalışan karar verici yapı olarak tanımlanmıştır. Ajanın temel amacı, yangın ve yoğun duman bulunan bölgelerden mümkün olduğunca kaçınarak en kısa ve en güvenli şekilde çıkış noktasına ulaşmaktır. Eğitim süreci boyunca ajan, gerçekleştirdiği hareketlerin sonuçlarına göre çevreden geri bildirim almakta ve bu geri bildirimler doğrultusunda öğrenme gerçekleştirmektedir.

Ajanın durum uzayı; çevrede bulunan duvar, ateş, duman, boş alan ve çıkış bilgilerini içermektedir. Durum, ajanın ızgara üzerindeki konumunu ve bulunduğu hücrenin tipini içeren s = (x, y, c) ikilisi olarak temsil edilmektedir. Burada x ve y koordinatları ajanın grid üzerindeki konumunu, c ∈ {boş, duvar, ateş, duman, çıkış} ise hücre tipini ifade etmektedir. Her hücre, ajanın karar verme sürecinde değerlendirdiği çevresel bir durumu temsil etmektedir. Aksiyon uzayı ise yukarı, aşağı, sağ ve sol olmak üzere dört temel hareketten oluşmaktadır. Ajan, her zaman adımında mevcut durumu değerlendirerek bu aksiyonlardan birini seçmekte ve yeni bir duruma geçiş yapmaktadır.

Ajanın hareket sistemi, pekiştirmeli öğrenme algoritmalarına uygun şekilde tasarlanmıştır. Gerçekleştirilen her hareket sonrasında ortam yeniden güncellenmekte ve yangın ile duman yayılımı tekrar hesaplanmaktadır. Böylece ajan, sürekli değişen çevre koşulları altında karar vermek zorunda kalmaktadır. Bu yapı sayesinde sistem, gerçek hayattaki acil durum senaryolarına daha yakın bir öğrenme ortamı sunmaktadır.

D. Ödül-Ceza Mekanizması

Pekiştirmeli öğrenme tabanlı sistemlerde ajanın doğru davranışları öğrenebilmesi için uygun bir ödül-ceza mekanizmasının oluşturulması büyük önem taşımaktadır. Bu çalışmada geliştirilen sistemde ajan, gerçekleştirdiği hareketlere göre farklı ödül ve ceza değerleri almaktadır.

Amaç, ajanın yalnızca en kısa yolu öğrenmesi değil, aynı zamanda güvenli rotaları tercih etmesini sağlamaktır.

Çıkış noktasına ulaşılması durumunda ajan yüksek pozitif ödül ile desteklenirken, yangın bölgelerine girilmesi durumunda yüksek negatif ceza uygulanmıştır. Duman bölgeleri ise insan sağlığı açısından risk oluşturduğu için orta seviyede negatif ödül ile değerlendirilmiştir. Ayrıca ajanın gereksiz hareketler yapmasını engellemek amacıyla her adım için küçük bir hareket maliyeti uygulanmıştır. Duvar hücrelerine hareket etmeye çalışılması durumunda ise negatif ceza verilmiştir.

Ödül-ceza mekanizmasında kullanılan değerler Denklem 4'de gösterilmektedir.

$$R(s') = \begin{cases} 500, & eğer\ s' çıkış\ noktası\ ise \\ -250, & eğer\ s' ölüm\ durumu\ ise \\ -20, & eğer\ s' ateş\ bölgesi\ ise \\ -5, & eğer\ s' duman\ bölgesi\ ise \\ -5, & eğer\ s' duvara\ çarpma\ durumu\ ise \\ -1, & diğer\ tüm\ normal\ hareketler\ için \end{cases} \quad (4)$$

Çıkış ödülünün yüksek seçilmesinin temel amacı, ajanın ana hedefe ulaşmayı öncelikli davranış olarak öğrenmesini sağlamaktır. Ateş bölgeleri insan yaşamı açısından yüksek risk taşıdığı için en büyük negatif ceza bu alanlara uygulanmıştır. Duman bölgeleri ise doğrudan ölümcül olmamakla birlikte görüş kaybı ve solunum problemleri oluşturduğu için orta seviyede ceza ile değerlendirilmiştir. Normal hareketlerde uygulanan düşük negatif ödül sayesinde ajanın gereksiz dolaşması engellenmiş ve daha kısa ve güvenli rotalara yönelmesi amaçlanmıştır.

E. Keşif ve Sömürü Stratejisi

Pekiştirmeli öğrenme algoritmalarında en önemli problemlerden biri, keşif (exploration) ve sömürü (exploitation) dengesinin doğru şekilde kurulmasıdır. Keşif aşamasında ajan, çevreyi daha iyi öğrenebilmek amacıyla farklı hareketleri denemektedir. Sömürü aşamasında ise daha önce öğrenilmiş en iyi hareketler tercih edilmektedir. Bu iki yaklaşım arasındaki dengenin doğru kurulması, öğrenme performansı açısından kritik öneme sahiptir.

Bu çalışmada epsilon-greedy keşif-sömürü stratejisi kullanılmıştır. Başlangıç epsilon değeri 1 olarak belirlenmiş ve her episode sonunda epsilon değeri 0.9997 katsayısı ile çarpılarak azaltılmıştır. Epsilon değerinin aşırı düşmesini engellemek amacıyla minimum epsilon değeri 0.05 olarak sınırlandırılmıştır. Böylece ajan başlangıç aşamasında çevreyi geniş ölçüde keşfederken, ilerleyen aşamalarda öğrendiği güvenli politikaları daha fazla kullanmaya başlamıştır.

Üstel epsilon azaltımı sayesinde ajan, başlangıç aşamasında farklı kaçış yollarını öğrenebilmiş; eğitim sürecinin ilerleyen bölümlerinde ise güvenli ve optimum rotalara odaklanmıştır. Bu yaklaşım, özellikle dinamik yangın ortamlarında daha kararlı ve güvenli davranışların öğrenilmesine katkı sağlamıştır.

IV. Benzetim Çalışmaları ve Performans Analizi

Eğitim süreci boyunca ajanlar, yangın ve dumanın dinamik olarak değiştiği grid tabanlı ortam içerisinde farklı senaryolar ile karşı karşıya bırakılmıştır. Amaç, ajanların yalnızca en kısa rotayı değil, aynı zamanda güvenli kaçış yollarını öğrenmesini sağlamaktır.

Eğitim sırasında ajanlar çevreden aldıkları ödül-ceza geri bildirimleri doğrultusunda öğrenme gerçekleştirmiştir. Başlangıç aşamasında rastgele hareket eden ajanlar, eğitim sürecinin ilerleyen bölümlerinde güvenli rotaları daha başarılı şekilde öğrenmeye başlamıştır. Özellikle epsilon-greedy tabanlı keşif-sömürü stratejisi sayesinde ajanların hem çevreyi keşfetmesi hem de öğrendikleri politikaları kullanması sağlanmıştır.

Geliştirilen sistemin eğitim sürecinde kullanılan temel parametreler Tablo I'de gösterilmektedir. Parametre değerleri belirlenirken ajanların öğrenme kararlılığı, eğitim süresi ve güvenli rota oluşturma performansı dikkate alınmıştır.

Tablo. I. Eğitim Parametreleri

Parametre	Değer
Episode Sayısı	30000
Öğrenme Katsayısı (α)	0.001
İndirim Faktörü (γ)	0.99
Başlangıç Epsilon Değeri	1
Final Epsilon Değeri	0.05
Epsilon Decay	0.9997
Maksimum Adım Sayısı	200

Öğrenme katsayısı (α), ajanın yeni deneyimlerden ne kadar hızlı öğrenebileceğini belirlemektedir. İndirim faktörü (γ) ise gelecekte elde edilecek ödüllerin öğrenme sürecindeki etkisini ifade etmektedir.

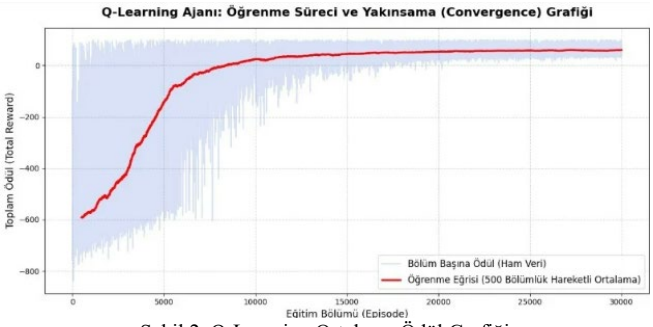
Eğitim süreci boyunca ajanlar, başlangıçta çevre hakkında herhangi bir bilgiye sahip olmadan rastgele hareketler gerçekleştirmiştir. İlk bölümlerde ajanların sık sık yangın bölgelerine girdiği, dumanlı alanlarda uzun süre kaldığı ve çıkış noktasına ulaşmakta başarısız olduğu gözlemlenmiştir. Ancak eğitim ilerledikçe ajanlar, çevreden aldıkları ödül-ceza geri bildirimlerini kullanarak daha güvenli hareket stratejileri geliştirmeye başlamıştır.

Eğitim sürecinin orta aşamalarında ajanların özellikle ateş bölgelerinden kaçınma davranışlarını öğrendiği ve daha güvenli koridorları tercih etmeye başladığı gözlemlenmiştir. Duman yayılımının artmasıyla birlikte ajanlar yalnızca yangından değil, aynı zamanda yoğun duman bulunan bölgelerden de uzak durmaya çalışmıştır. Bu durum, ödül-ceza mekanizmasının güvenlik odaklı öğrenme üzerinde etkili olduğunu göstermektedir.

Eğitim sürecinin son aşamalarında ise ajanların çıkış noktalarına daha kısa sürede ulaştığı ve başarısız tahliye oranlarının önemli ölçüde azaldığı gözlemlenmiştir. Özellikle üstel epsilon azaltımı sayesinde ajanlar, başlangıçta çevreyi geniş ölçüde keşfetmiş; ilerleyen aşamalarda ise öğrendikleri güvenli politikaları daha etkin şekilde kullanmıştır.

Eğitim sürecinin başlangıç aşamalarında ajanların büyük bölümü çevre hakkında yeterli bilgiye sahip olmadığı için rastgele hareketler gerçekleştirmiştir. Bu nedenle ilk bölümlerde toplam ödül değerlerinin yaklaşık -600 seviyelerinde olduğu gözlemlenmiştir. Eğitim ilerledikçe ajanların çevreden aldıkları ödül-ceza geri bildirimlerini kullanarak daha güvenli hareket stratejileri geliştirdiği belirlenmiştir.

Şekil 2 incelendiğinde, Q-Learning algoritmasının yaklaşık 10000. bölümden sonra kararlı öğrenme davranışı göstermeye başladığı görülmektedir. Eğitim sürecinin ilerleyen aşamalarında toplam ödül değerleri pozitif seviyelere ulaşmış ve yaklaşık +60 seviyesinde kararlı hale gelmiştir. Bu durum, ajanın güvenli çıkış rotalarını başarılı şekilde öğrenebildiğini göstermektedir.



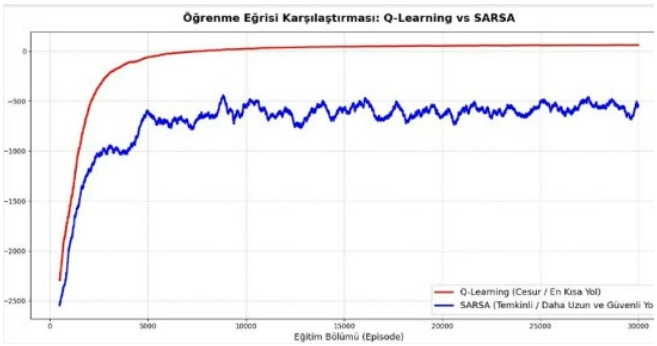
Şekil.2. Q-Learning Ortalama Ödül Grafiği

Q-Learning ve SARSA algoritmalarının öğrenme performansları karşılaştırıldığında, Q-Learning algoritmasının daha hızlı yakınsama gerçekleştirdiği gözlemlenmiştir. Şekil 3'de görüldüğü üzere Q-Learning algoritması yaklaşık 10000. bölüm civarında kararlı öğrenme davranışı sergilerken, SARSA algoritmasının yakınsama sonrası ortalama ödül değeri yaklaşık -550 seviyesinde kalırken öğrenme eğrisindeki dalgalanma devam etmiştir.

SARSA algoritmasının toplam ödül değerleri yaklaşık -500 ila -700 aralığında değişim göstermiştir. Bunun temel nedeni, SARSA algoritmasının riskli bölgelerden daha fazla kaçınması ve daha temkinli hareket stratejileri geliştirmesidir. Bu durum SARSA algoritmasının bazı senaryolarda tahliye süresinin uzamasına ve dinamik tehlikelere daha uzun süre maruz kalınmasına neden olmuştur. Buna karşılık Q-Learning algoritması, maksimum ödülü hedefleyen yapısı nedeniyle daha kısa ancak bazı durumlarda daha riskli rotaları tercih etmiştir. Tablo II'te her iki algoritmanın performans karşılaştırması verilmiştir.

Tablo. II. Q-Learning SARSA Performans Karşılaştırılması

Metrik	Q-Learning	SARSA
Yakınsama Bölümü (Episode)	~ 10000	~ 12000
Ortalama Toplam Ödül	61.94	-283.84
Ortalama Adım Sayısı	39.06	371.21
Başarı Oranı	%99.45	%86.5



Şekil.3. Q-Learning SARSA Karşılaştırma Grafiği

V. Sonuçlar ve Gelecek Çalışmalar

Bu bölümde Q-Learning ve SARSA algoritmalarına ilişkin eğitim sonuçları değerlendirilmektedir. Her iki ajan da 30.000 bölüm boyunca yangın ve dumanın dinamik biçimde yayıldığı grid ortamında eğitilmiş; performans ölçütü olarak bölüm başına toplam ödül, çıkışa ulaşma oranı ve riskli bölgelere girme sıklığı izlenmiştir. Simülasyonlar, geliştirilen sistemin dinamik yangın ortamlarında geçerli tahliye rotaları ürettiğini ortaya koymaktadır. Üstel epsilon azaltımı, komşu hücre etkileşimli yangın yayılım modeli ve güvenlik ağırlıklı ödül yapısının bir arada kullanılması, ajanların tehlikeli bölgelerden uzak duran politikalar öğrenmesine doğrudan katkı sağlamıştır. İki algoritma karşılaştırıldığında, Q-Learning daha hızlı yakınsama sergilerken SARSA riskli bölgelerden daha tutarlı biçimde kaçınan politikalar üretmiştir. Bu bulgu, insan güvenliğinin öncelikli olduğu acil durum uygulamalarında sırf performans odaklı algoritmaların ötesine geçilerek güvenlik güdümlü öğrenme yaklaşımlarının benimsenmesi gerektiğini göstermektedir.

Geliştirilen sistemin, gelecekte gerçek zamanlı sensör verileri ile desteklenerek akıllı bina teknolojileri ve afet yönetim sistemlerinde kullanılabilecek potansiyeline sahip olduğu değerlendirilmektedir. Geliştirilen sistem grid tabanlı simülasyonda tatmin edici sonuçlar verse de gerçek yangın ortamları çok daha gürültülü, çok katmanlı veriler ve öngörülemeyen insan hareketleri içermektedir. Bu sınırlamaları aşmak ve sistemi gerçek kullanım koşullarına yaklaştırmak amacıyla bir dizi ek çalışma planlanmaktadır.

İlerleyen çalışmalarda sıcaklık sensörleri, duman sensörleri ve nesnelerin interneti (IoT) tabanlı akıllı bina sistemlerinden alınacak gerçek zamanlı verilerin modele entegre edilmesi planlanmaktadır. Böylece sistemin gerçek bina ortamlarında anlık çevresel değişimlere göre dinamik tahliye yönlendirmesi yapabilmesi amaçlanmaktadır. Ayrıca mevcut çalışmada kullanılan Q-Learning ve SARSA algoritmalarına ek olarak Deep Q-Network (DQN) tabanlı derin pekiştirmeli öğrenme yönteminin kullanılması hedeflenmektedir. Bu sayede daha büyük durum uzaylarında ve karmaşık bina yapılarında daha başarılı öğrenme performansı elde edilmesi amaçlanmaktadır. Bunun yanında gelecekte çoklu ajan sistemleri kullanılarak kalabalık insan davranışlarının modellenmesi, panik etkisinin simülasyona dahil edilmesi ve bireyler arası etkileşimlerin tahliye performansına etkisinin incelenmesi planlanmaktadır. Gerçek bina haritaları ve üç boyutlu ortam modelleri kullanılarak sistemin farklı afet senaryoları altında test edilmesi de hedeflenen çalışmalar arasındadır.

Kaynakça

[1] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, pp. 4282–4286, 1995.

[2] C. Burstedde, K. Klauck, A. Schadschneider, and J. Zittartz, "Simulation of pedestrian dynamics using a two-dimensional cellular automaton," *Physica A: Statistical Mechanics and its Applications*, vol. 295, no. 3-4, pp. 507-525, 2001.

[3] R. Sutton and A. Barto, Reinforcement Learning: An Introduction, 2018.

[4] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279-292, 1992.

[5] G. A. Rummery and M. Niranjan, "On-line Q-learning using connectionist systems," *Cambridge University Engineering Department*, Tech. Rep. CUED/F-INFENG/TR 166, 1994.

[6] S. Gwynne, Edwin R. Galea, M. Owen, Peter Lawrence, L. Filippidis, “A review of the methodologies used in evacuation modelling,” 1999.

MODERN FİZİK SANAL LABORATUVARI: CODATA 2018 KALİBRASYONLU ON KLASİK KUANTUM DENEYİNİN WEB TABANLI İNTERAKTİF SİMÜLASYON ORTAMINDA GERÇEKLEŞTİRİLMESİ

Modern Physics Virtual Laboratory: Implementation of Ten Classical Quantum Experiments in a Web-Based Interactive Simulation Environment Calibrated with CODATA 2018

Cafer Mert Yeşilkanat[1*] ve Serkan Akkoyun[2,3]

[1] Artvin Çoruh Üniversitesi Matematik ve Fen Eğitimi Bölümü, Artvin, Türkiye

[2] Sivas Cumhuriyet Üniversitesi Fizik Bölümü, 58140, Sivas, Türkiye

[3] School of Mathematics and Physics, University of Surrey, Guildford, United Kingdom

[*]

E-posta: *cmyesilkanat@artvin.edu.tr*

ÖZET

Bu çalışmada, lisans düzeyinde modern fizik eğitimine yönelik web tabanlı bir sanal laboratuvar sistemi (*Modern Fizik Sanal Laboratuvarı*) tasarlanmış ve uygulamaya konulmuştur. HTML5 Canvas API, JavaScript ve KaTeX matematik işleme altyapısı üzerine inşa edilen sistem; fotoelektrik etki, karacisim ışıması, Franck-Hertz deneyi, Compton saçılması, Millikan yağ damlası deneyi, çift yarık deneyi (dalga-parçacık dualitesi), hidrojen spektrumu, Davisson-Germer deneyi, Rutherford saçılması ve Stern-Gerlach deneyi olmak üzere on klasik modern fizik deneyini barındırmaktadır. Sistemin ayırt edici özelliği, tüm hesaplamaların 2018 yılı CODATA önerilen değerleri (Planck sabiti h = 6.62607015×10^{-34} J·s, Boltzmann sabiti kB = 1.380649×10^{-23} J·K^{-1}, elementsel yük e = 1.602176634×10^{-19} C) ile kalibre edilmiş gerçek zamanlı fizik motoru tarafından yürütülmesidir. Grafik motoru, toplanan deney verilerine doğrusal regresyon (R^2 hesabı dahil) uygulayarak Planck sabiti, Wien sabiti, Compton dalga boyu, Rydberg sabiti ve Bohr magnetonu gibi temel sabitlerin tayin edilmesine olanak tanımaktadır. Her deneye eşlik eden altı bölümlü konu anlatımı modülü (tarih, teori, prosedür, sonuçlar, uygulamalar, hata analizi) sistemi yalnızca simülasyon değil, bütünleşik bir öğrenme ortamı haline getirmektedir. Sistem, herhangi bir kurulum gerektirmeksizin standart web tarayıcılarında 60 fps animasyon ile çalışmakta ve toplanan verilerin CSV formatında dışa aktarılmasını desteklemektedir.

Anahtar Kelimeler: *sanal laboratuvar, modern fizik, fotoelektrik etki, Planck sabiti, CODATA 2018, web tabanlı simülasyon, fizik eğitimi, HTML5 Canvas*

ABSTRACT

In this study, a web-based virtual laboratory system (*Modern Physics Virtual Laboratory*) was designed and implemented for undergraduate modern physics education. Built upon the HTML5 Canvas API, JavaScript, and the KaTeX mathematical rendering library, the system incorporates ten classical modern physics experiments: the photoelectric effect, blackbody radiation, the Franck-Hertz experiment, Compton scattering, the Millikan oil-drop experiment, the double-slit experiment (wave-particle duality), the hydrogen spectrum, the Davisson-Germer experiment, Rutherford scattering, and the Stern-Gerlach experiment. The distinguishing characteristic of the system is its real-time physics engine calibrated with the CODATA 2018 recommended values of fundamental constants (Planck constant $h = 6.62607015\times10^{-34}$ J·s, Boltzmann constant kB $= 1.380649\times10^{-23}$ J·K^{-1}, elementary charge $e = 1.602176634\times10^{-19}$ C). The built-in graph engine applies linear regression with R^2 calculation to collected data, enabling determination of fundamental constants such as Planck's constant, Wien's constant, the Compton wavelength, the Rydberg constant, and the Bohr magneton. A six-tab tutorial module (history, theory, procedure, results, applications, error analysis) accompanying each experiment transforms the system into a comprehensive learning environment rather than a mere simulation tool. The system operates at 60 fps animation in standard web browsers without any installation requirements and supports CSV export of collected data.

Keywords: *virtual laboratory, modern physics, photoelectric effect, Planck constant, CODATA 2018, web-based simulation, physics education, HTML5 Canvas*

1. GIRIŞ

Modern fizik, 20. yüzyılın başında klasik fiziğin açıklayamadığı deneysel gözlemler üzerine kurulmuştur. Fotoelektrik etkiyi açıklayan Einstein'ın foton hipotezi (1905), karacisim ışımasını formüle eden Planck'ın kuantum varsayımı (1900), elektronların dalga doğasını kanıtlayan Davisson-Germer deneyi (1927) ve atom çekirdeğinin varlığını ortaya koyan Rutherford saçılması (1911); modern fiziğin yalnızca birer tarihsel sayfa değil, çağdaş teknolojinin temel taşları olduğunu göstermektedir. Bu deneylerin lisans düzeyinde kavranması, ileri fizik, malzeme bilimi, yarı iletken teknolojisi ve kuantum bilişim gibi alanlarda yetkinlik için zorunlu bir ön koşuldur.

Ancak bununla beraber modern fizik deneylerinin gerçek laboratuvar ortamında yürütülmesi ciddi güçlükler barındırmaktadır. Millikan yağ damlası düzeneği gibi aparatlar hassas mekanik ayarlama ve uzun gözlem süreleri gerektirirken, Franck-Hertz düzeneği yüksek gerilim ve zehirli cıva buharı içermektedir. Compton saçılması ve Rutherford saçılması deneyleriyse radyoaktif kaynak kullanımını zorunlu kılmaktadır. Bu koşullarda tüm deneylere eşzamanlı erişim olanağı, çoğu üniversitede sınırlı kalmaktadır. Öte yandan coğrafi kısıtlamalar, uzaktan eğitim gereksinimleri ve

pandemi sürecinin ortaya koyduğu pratik zorunluluklar, sanal laboratuvar çözümlerine olan ilgiyi belirgin biçimde artırmıştır (Radhamani ve ark., 2021; Gamage ve ark., 2020).

Literatürde sanal fizik laboratuvarlarına ilişkin öncü çalışmalar incelendiğinde, PhET İnteraktif Simülasyonları projesinin (Colorado Üniversitesi) bu alandaki en kapsamlı girişimlerden biri olduğu görülmektedir. Wieman ve Perkins (2005), PhET simülasyonlarının öğrencilerin kavramsal anlamalarına katkısını ölçmüş; aktif öğrenme, öz keşif ve gerçek zamanlı geri bildirim mekanizmalarının geleneksel öğretime kıyasla anlamlı kazanımlar sağladığını bildirmiştir. Perkins ve arkadaşları (2006) ise bu simülasyonların farklı öğrenme stillerine uyum sağlayabildiğini ve hem lisede hem lisans düzeyinde etkin biçimde kullanılabildiğini göstermiştir. Bununla birlikte, mevcut platformların önemli bir bölümü belirli deneylere odaklanmakta; temel fizik sabitlerinin doğrudan tayin edilebildiği, gerçek zamanlı regresyon analizi sunan ve CODATA 2018 doğruluk standartlarıyla kalibre edilmiş entegre bir sistem sunmamaktadır.

Bu çalışma, söz konusu boşluğu doldurmaya yönelik olarak tasarlanan *Modern Fizik Sanal Laboratuvarı*'nı tanıtmaktadır. Söz konusu sistem; (1) tek bir web sayfasında on klasik modern fizik deneyini barındırması, (2) hesaplamaların CODATA 2018 sabitleriyle kalibre edilmiş gerçek zamanlı bir fizik motoru tarafından yürütülmesi, (3) kullanıcıların toplanan verilerden lineer regresyon yöntemiyle temel fizik sabitlerini bizzat tayin edebilmeleri, (4) her deneye eşlik eden kapsamlı konu anlatımı modülünün bulunması ve (5) herhangi bir eklenti ya da kurulum gerektirmeden standart web tarayıcılarında çalışması özellikleriyle literatürdeki benzer platformlardan ayrışmaktadır.

Makalenin geri kalan bölümleri şöyle düzenlenmiştir: Bölüm 2'de sistemin tasarımı, mimarisi ve uygulamaya konulan deney modülleri açıklanmaktadır. Bölüm 3'te her deneyin simülasyon mantığı, ölçüm yöntemi ve elde edilebilecek fizik sabitleri tartışılmaktadır. Bölüm 4'te genel değerlendirme ve eğitim potansiyeli ele alınmakta, Bölüm 5'te ise sonuçlar sunulmaktadır.

2. MATERYAL VE YÖNTEM

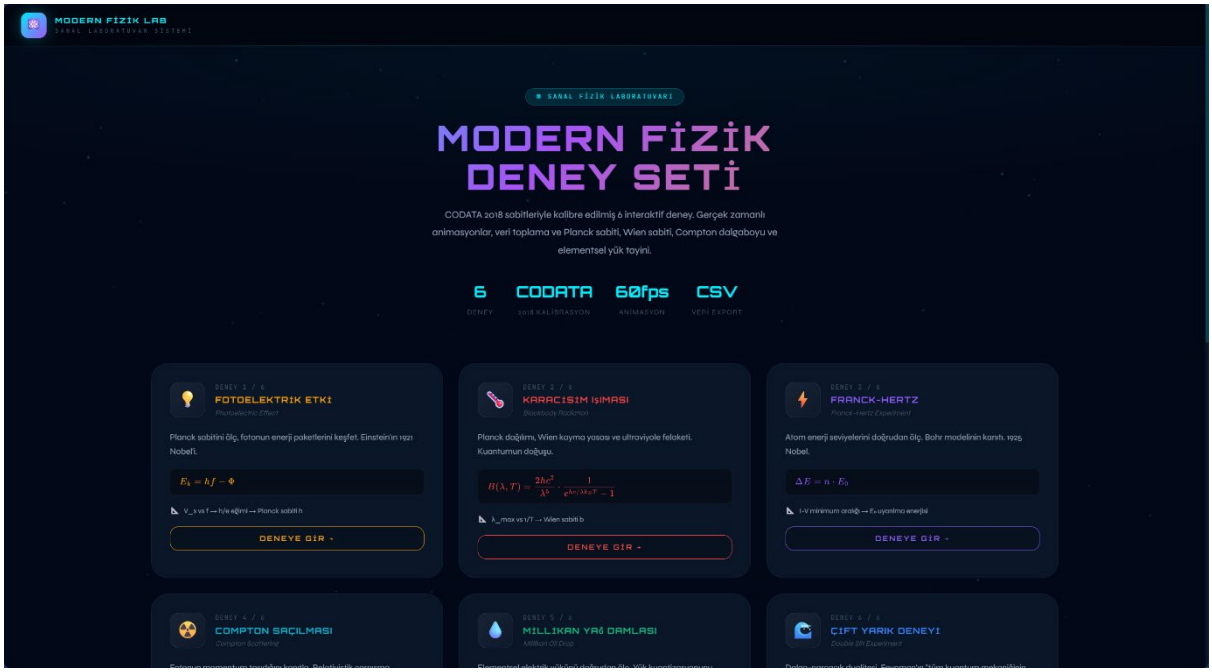
2.1. Sistem Mimarisi ve Kullanılan Teknolojiler

Modern Fizik Sanal Laboratuvarı, tarayıcı tabanlı (client-side) mimaride geliştirilmiş tek sayfalık bir web uygulamasıdır (Single Page Application, SPA). Sunucu tarafında herhangi bir altyapı gerektirmeyen bu yaklaşım, sistemin çevrimdışı ortamda da çalışabilmesine ve kurulum gerektirmeksizin dağıtılabilmesine olanak tanımaktadır. Uygulamanın temel teknolojik bileşenleri şunlardır:

- HTML5 Canvas API: Her deneyin gerçek zamanlı animasyonunun işlendiği 2B grafik arayüzü. requestAnimationFrame döngüsüyle 60 fps akıcılık hedeflenmektedir.

- Vanilla JavaScript (ES2020): Nesne yönelimli programlama paradigmasıyla tasarlanmış fizik motoru, grafik motoru ve deney sınıfları.
- KaTeX v0.16.9: Matematiksel formüllerin LaTeX söz diziminden tarayıcı içi tipografik kalitede işlenmesi için kullanılan açık kaynaklı kütüphane.
- CSS3 Özel Değişkenleri (Custom Properties): Her deneye atanmış renk şemasının dinamik olarak güncellenebildiği tematik stil sistemi.
- Google Fonts: Orbitron (başlık/display), Syne (gövde metni) ve JetBrains Mono (sayısal değerler/kod) olmak üzere üç tipografik katman.

Uygulama, iki ana görünümden oluşmaktadır: (a) tüm deneylerin kart formatında listelendiği Ana Panel (*dashboard*) ve (b) seçilen deneyin tam ekran düzeninde çalıştırıldığı Deney Görünümü (*experiment view*). Deney görünümü; sol tarafta simülasyon tuvalini, sağ tarafta kontrol panelini (kaydırıcılar, seçim kutuları, LCD ölçüm ekranları ve veri toplama araçları) ve alt kısımda gerçek zamanlı veri grafiğini içeren üç bölümlü bir yerleşim şemasına (*grid layout*) sahiptir. Duyarlı tasarım (responsive design) ilkesiyle oluşturulan sistem, 1080 piksel genişliğin altındaki ekranlarda dikey yığılma düzenine geçmektedir.



Şekil 1: *Uygulamanın ana paneli: 10 deney kartı, her birinde temel formül, ölçüm hedefi ve erişim butonu görünmektedir. Arka planda parçacık animasyonu çalışmaktadır.*

2.2. Fizik Motoru: CODATA 2018 Sabit Seti

Sistemin doğruluk altyapısı, *PHY* adlı merkezi sabit nesnesi üzerine kuruludur. Tüm deney hesaplamaları bu nesneden alınan değerleri kullanmaktadır; herhangi bir sabit doğrudan kod içine gömülmemiştir. Kullanılan CODATA 2018 değerleri Tablo 1'de verilmektedir.

Tablo 1: Sistemde kullanılan CODATA 2018 temel fizik sabitleri (Tiesinga et al., 2021).

Sabit	Sembol / Değer	Birim	Kullanıldığı Deney
Planck Sabiti	$h = 6.62607015\times10^{-34}$	J·s	Foto, BB, Compton, Davisson
İndirgenmiş Planck	$\hbar = 1.054571817\times10^{-34}$	J·s	Hesaplama
Işık Hızı	$c = 2.99792458\times10^{8}$	m·s⁻¹	Tüm deneyler
Elementsel Yük	$e = 1.602176634\times10^{-19}$	C	Foto, Millikan, FH, SG
Elektron Kütlesi	$m_e = 9.1093837015\times10^{-31}$	kg	Compton, Davisson
Boltzmann Sabiti	$k_B = 1.380649\times10^{-23}$	J·K⁻¹	Karacisim, Stern-Gerlach
Stefan-Boltzmann	$\sigma = 5.670374419\times10^{-8}$	W·m⁻²·K⁻⁴	Karacisim
Wien Sabiti	$b = 2.897771955\times10^{-3}$	m·K	Karacisim
Compton Dalga Boyu	$\lambda_C = 2.42631023867\times10^{-12}$	m	Compton
Avogadro Sayısı	$N_A = 6.02214076\times10^{23}$	mol⁻¹	Referans

2.3. Grafik Motoru ve İstatistiksel Analiz

Her deney modülüne entegre edilmiş *GraphEngine* sınıfı, Canvas 2D bağlamı üzerinde gerçek zamanlı grafik çizmektedir. Motor şu özelliklere sahiptir:

- Milimetrik kâğıt arka planı: 5 piksel aralıklı ince yardımcı çizgiler ve 25 piksel aralıklı kalın çizgilerden oluşan ızgara yapısı.
- Etkileşimli fare ipucu (tooltip): İmleç bir veri noktasına yaklaştığında o noktanın koordinatları (x, y) ve birimleri anlık olarak görüntülenmektedir.
- Doğrusal regresyon ve R² hesabı: En az iki nokta toplandığında en küçük kareler yöntemiyle eğim (m), y-kesim noktası (b) ve determinasyon katsayısı (R^2) otomatik hesaplanarak grafik üzerinde görüntülenmektedir.
- Teorik eğri katmanı: Mevcut parametrelere göre hesaplanan teorik eğri, kesik çizgi formatında gerçek zamanlı olarak çizilmektedir.
- CSV dışa aktarma: Toplanan veri çiftleri başlık satırıyla birlikte standart CSV formatında indirilebilmektedir.

Doğrusal regresyon algoritması, N adet (x_i, y_i) veri noktası için aşağıdaki ifadeleri kullanmaktadır: Burada m eğim; n, y- kesim/intercept ve R^2 determinasyon katsayısıdır.

$$m = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{N\sum x_i^2 - (\sum x_i)^2}$$

$$n = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{N\sum x_i^2 - (\sum x_i)^2}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Bu doğrusal regresyon modelinde kullanılan parametrelerin ve istatistiksel operatörlerin tanımları şu şekildedir:

- N: Veri setindeki toplam gözlem (ölçüm noktası) sayısı.
- x_i : i. gözleme ait bağımsız değişken (girdi) değeridir.
- y_i : i. gözleme ait bağımlı değişken (ölçülen çıktı) değeridir.
- $\hat{y}_i$: i. veri noktası için model tarafından hesaplanan tahmini (fit edilmiş) değerdir.
- $\bar{y}$: Tüm gözlenen y değerlerinin aritmetik ortalamasıdır.

Bu istatistiksel altyapı, öğrencilerin hem simülasyon verilerinden temel fizik sabitlerini bizzat çıkarmasına olanak tanımakta hem de deneysel fizik metodolojisinin (veri toplama, grafik, regresyon ve sabit tayini) sanal ortamda yeniden üretilmesini mümkün kılmaktadır.

2.4. Konu Anlatımı Modülü

Her deneye eşlik eden konu anlatımı modülü, altı sekmeli bir modal pencere arayüzü sunmaktadır. Sekme yapısı Tablo 2'de özetlenmektedir.

Tablo 2: Her deneyde bulunan konu anlatımı modülünün sekme yapısı.

No	Sekme Başlığı	İçerik
1	Tarih	Zaman çizelgesi formatında deneyin keşif süreci, öncü bilim insanlarının katkıları ve Nobel ödülleri.
2	Teori	KaTeX ile işlenmiş matematiksel formüller, türetmeler ve temel kavramların açıklanması.
3	Prosedür	Adım adım deney protokolü: hangi parametrenin değiştirileceği, hangi verinin kaydedileceği, nasıl analiz edileceği.
4	Sonuçlar	Beklenen grafik çıktıları, tayin edilecek sabitler ve kabul edilebilir hata aralıkları.
5	Uygulamalar	Deneyin günlük teknoloji ve araştırma alanlarındaki pratik yansımaları.
6	Hata Analizi	Sistematik ve rastgele hata kaynakları, bağıl hata formülleri ve sonuç doğrulama kriterleri.

2.5. Deney Modülleri: Genel Tasarım Prensibi

Her deney, *BaseExp* temel sınıfından türetilen bağımsız bir JavaScript sınıfı olarak uygulanmıştır. Bu sınıflar; *mount(container)*, *stop()* ve *buildHTML()* arayüzünü paylaşmaktadır. Bu tasarım sayesinde kullanıcı ana panel ile deney görünümü arasında geçiş yaptığında bir önceki deneyin animasyon döngüsü temiz biçimde durdurulmakta ve bellek sızıntısı önlenmektedir. Her deney sınıfının beş temel bileşeni bulunmaktadır: (1) simülasyon parametrelerinin tutulduğu durum nesnesi, (2) HTML arayüzünün üretildiği *buildHTML()* metodu, (3) fizik hesaplamalarının yürütüldüğü ve Canvas'a çizim

yapıldığı animasyon döngüsü, (4) kullanıcı etkileşimlerini işleyen olay dinleyicileri, (5) veri toplama ve CSV dışa aktarma metodları.

3. BULGULAR VE TARTIŞMA

Bu bölümde sisteme dahil edilen on deney modülü fizik içeriği, simülasyon tasarımı, ölçüm metodolojisi ve beklenen sonuçlar açısından ayrı alt bölümler halinde ele alınmaktadır. Her modül; simülasyon görünümü, kontrol parametreleri, veri toplama protokolü ve tayin edilecek fizik sabiti bakımından değerlendirilmektedir.

3.1. Deney Modülleri

3.1.1 Fotoelektrik Etki Deneyi

Fotoelektrik etki, ışığın kuantum doğasının deneysel kanıtı olup Einstein'ın 1905 tarihli foton hipotezine dayanmaktadır (Einstein, 1905). Millikan'ın 1914 yılında başlattığı ve aslında Einstein'ı yanlışlamayı hedeflediği ölçümler, tam tersine foton modelini olağanüstü hassasiyetle doğrulamış ve Planck sabitinin deneysel tayinine zemin hazırlamıştır (Millikan, 1916).

Simülasyonda sekiz farklı metal (Sodyum, Potasyum, Sezyum, Çinko, Bakır, Alüminyum, Gümüş, Platin) iş işlevi değerleriyle birlikte seçilebilmektedir. Kullanıcı, ışık frekansını 3.0×10^{14} Hz (kırmızı/IR) ile 15.0×10^{14} Hz (derin UV) arasında, ışık şiddetini ise %0 ile %100 arasında sürekli değiştirebilmektedir. Simülasyon tuvali, foton parçacıklarının metal yüzeyine çarpmasını ve yeterli enerjiye ulaşıldığında elektron saçılmasını animasyonlu olarak göstermektedir. Dört LCD ekran; foton enerjisi ($E = hf$), kinetik enerji ($Ek = hf - \Phi$), durdurucu gerilim (Vs) ve fotoakım değerlerini anlık olarak görüntülemektedir.

Deney protokolü; seçilen metal için 6–8 farklı frekans değerinde Vs ölçümünün kaydedilmesini, ardından $Vs - f$ grafiğinin regresyon analiziyle değerlendirilmesini içermektedir.

Doğru uygulanan bir ölçüm protokolünde $R^2 > 0.998$ ve Planck sabitinde %1'den az bağıl hata beklenmektedir. Önemli bir kavramsal vurgu: ışık şiddeti artırıldığında Vs değişmemekte, yalnızca fotoakım artmaktadır. Bu durum, dalga teorisiyle bağdaşmayan ve foton modelinin en güçlü kanıtı olan deneysel gerçeği öğrencinin bizzat gözlemlemesine imkân tanımaktadır.

3.1.2. Karacisim Işıması Deneyi

Karacisim ışıması problemi, 19. yüzyıl sonunda klasik elektromanyetizmanın karşılaştığı en derin çelişkiyi temsil etmektedir. Rayleigh-Jeans yasası kısa dalga boylarında enerji yoğunluğunun sonsuza yaklaştığını öngörürken, deneyler saçılan enerjinin belirli bir dalga boyunda maksimuma ulaşıp azaldığını göstermektedir. Bu çelişki "ultraviyole felaketi" adıyla tarihe geçmiştir. Planck, 1900 yılında enerji değişiminin $\varepsilon = hf$ büyüklüğündeki kesikli paketler biçiminde gerçekleştiğini varsayarak karacisim spektrumunu başarıyla modelleyen dağılım yasasını elde etmiştir (Planck, 1901).

Simülasyon, 300 K ile 30000 K arasında ayarlanabilen sıcaklık değerlerine karşılık gerçek zamanlı spektral yoğunluk eğrisi çizmektedir. Siyah cisim rengi, sıcaklığa bağlı renk algoritmasıyla (kırmızı, turuncu, sarı, beyaz ve mavi-beyaz) görselleştirilmektedir. Kullanıcı Rayleigh-Jeans eğrisini açarak klasik tahminle kuantum teorisinin diverjansını doğrudan karşılaştırabilmektedir. $\lambda max - 1/T$ grafiği, Wien kayma yasasının $b = \lambda max \cdot T$ sabitini ortaya koymaktadır.

Güneş (5778 K, λmax ≈ 502 nm), oda sıcaklığı (300 K, λmax ≈ 9700 nm, kızılötesi) ve mavi yıldız (30000 K, λmax ≈ 97 nm, UV) gibi gerçek astrofizik nesnelere karşılık gelen sıcaklıklar girilerek Wien yasasının evrensel geçerliliği test edilebilmektedir. Bu özellik, deneyi astrofizik ve termal fiziğin birleşim noktasında konumlandırmaktadır.

3.1.3. Franck-Hertz Deneyi

James Franck ve Gustav Hertz, 1914 yılında yürüttükleri deney ile Bohr atom modelinin öngördüğü enerji seviyelerinin varlığını doğrudan kanıtlamış; bu çalışma 1925 Nobel Fizik Ödülü'ne layık görülmüştür (Franck & Hertz, 1914). Deney; katottan salınan elektronların hızlandırıcı gerilimle ivmelendirilerek cıva (Hg) ya da neon (Ne) gazı içinden geçirilmesi ve anot akımının bu gerilime göre incelenmesi üzerine kuruludur. Elektronların kinetik enerjisi E_0 uyarılma enerjisine ulaştığında, kinetik enerji gazın atomlarına aktarılarak inelastik çarpışma gerçekleşmekte ve anot akımı azalmaktadır.

Simülasyonda iki gaz seçeneği bulunmaktadır: Cıva (E_0 = 4.89 eV, λ = 253.7 nm) ve Neon (E_0 = 18.7 eV, λ = 585.2 nm). Hızlandırıcı gerilim 0–80 V arasında ve fırın sıcaklığı 300–800 K arasında ayarlanabilmektedir. $I–V$ eğrisi, periyodik minimumlara sahip karakteristik dalgalı yapıyı sergilemelidir. Minimum aralığı ölçümünden uyarılma enerjisi E_0 direkt olarak tayin edilmektedir:

Sıcaklık değişiminin (termal dağılımın) minimum derinliği üzerindeki etkisi, gerçek laboratuvar koşullarında gözlemlenen yumuşama davranışıyla tutarlı biçimde modellenmiştir. Bu özellik, deneyin hem temel kuantum mekaniği içeriğini hem de gerçekçi deneysel koşulları karşılaştırma fırsatı sunmaktadır.

3.1.4. Compton Saçılması Deneyi

Arthur Holly Compton, 1923 yılında X-ışınlarının hafif elementler tarafından saçılmasında dalga boyunun arttığını gözlemlemiş ve bu olayın foton-elektron relativistik çarpışması olarak modellenebildiğini göstermiştir (Compton, 1923). Bu çalışma, fotonun hem enerji hem de momentum taşıdığının doğrudan kanıtı olarak kabul edilmekte olup Compton'a 1927 Nobel Fizik Ödülü'nü kazandırmıştır.

Simülasyon üç *X*-ışını kaynağını desteklemektedir: Mo Kα (17.4 keV, λ_0 = 0.0712 nm), Ag Kα (22.1 keV, λ_0 = 0.0559 nm) ve Cs-137 γ (661.7 keV, λ_0 = 0.00187 nm). Saçılma açısı θ, 0°–180° arasında değiştirilebilmekte; LCD panelleri saçılmış foton enerjisi, geri tepme elektronu enerjisi ve Δλ değerlerini anlık olarak güncellemektedir.

$\Delta\lambda - (1 - \cos\theta)$ grafiğinin regresyon eğimi, Compton dalga boyunu $\lambda_C = 2.426\times10^{-12}$ m olarak vermektedir. Bu değer CODATA 2018 sabitiyle ($\lambda_C = 2.42631023867\times10^{-12}$ m) karşılaştırıldığında beklenen sapmanın %1'in altında kalması gerekmektedir.

3.1.5. Millikan Yağ Damlası Deneyi

Robert A. Millikan, 1909–1913 yılları arasında yürüttüğü yağ damlası deneyiyle elementsel elektrik yükünü doğrudan ölçmeyi başarmış ve yük kuantizasyonunu kanıtlamıştır. Bu çalışma 1923 Nobel Fizik Ödülü'nü beraberinde getirmiştir. Deneyin temel denklemi, damlaya etkiyen elektrik kuvveti, yerçekimi kuvveti ve viskoz sürükleme kuvvetleri arasındaki dengeye dayanmaktadır (Millikan, 1913):

Simülasyonda 1–12 adet yük taşıyan damlalar oluşturulabilmekte, elektrik alan şiddeti 0–200 kV/m arasında ayarlanabilmektedir. Damlanın serbest düşme hızından yarıçap, ardından yükselme hızından yük değeri hesaplanmaktadır. Toplanan q değerlerinin en büyük ortak bölen analizi ile elementsel yük e tayin edilmektedir. Bu analiz yöntemi, yük kuantizasyonunu sayısal olarak göstermesi bakımından pedagojik açıdan özellikle değerlidir.

3.1.6. Çift Yarık Deneyi (Dalga-Parçacık Dualitesi)

Richard Feynman, çift yarık deneyini tüm kuantum mekaniğinin özünü taşıyan bir deney olarak nitelendirmiştir (Feynman 1965). Young'ın 1801'de ışık için gerçekleştirdiği bu deney, daha sonra tek tek fotonlar ve hatta elektronlar için yinelenmiş; her seferinde aynı girişim deseni gözlemlenmiştir. Deney, dalga-parçacık ikileminin en saf gösterimidir.

Simülasyon hem foton, hem de elektron modunda çalışabilmektedir. Foton modunda dalga boyu 380–700 nm (görünür spektrum) arasında seçilebilmekte; saçak aralığı $\Delta y = \lambda L/d$ formülüyle hesaplanmaktadır. Elektron modunda ise hızlandırma geriliminden de Broglie dalga boyu $\lambda dB = h/\sqrt{2\, m_e\, e\, V}$ hesaplanmakta ve elektronların girişim deseni oluşturduğu gösterilmektedir. Yarık genişliği a ve yarık aralığı d bağımsız olarak değiştirilebilmektedir. Ayrıca saçak aralığı ölçümlerinden dalga boyu tayini yapılabilmekte; elektron modu de Broglie hipotezinin (de Broglie, 1924) görsel kanıtını sunmaktadır.

3.1.7. Hidrojen Spektrumu Deneyi

Hidrojen atomunun çizgi spektrumu, Bohr atom modelinin (1913) en güçlü kanıtlarından biridir. Balmer (1885), görünür bölgedeki çizgileri ampirik bir formülle tanımlamış; Rydberg ise daha kapsamlı bir formül geliştirmiştir. Bohr, bu emprik yasanın kuantum koşullarından türetilebileceğini göstermiştir (Bohr, 1913).

Simülasyon, $n_1 = 1–4$ ve $n_2 = n_1+1$ ile n_1+10 arasındaki tüm geçişler için dalga boyunu ve karşılık gelen rengi gerçek zamanlı hesaplamaktadır. Dalga boyuna karşılık gelen görünür renk, Balmer

serisi için spektral renk algoritmasıyla gösterilmektedir. $1/\lambda - (1/n_1^2 - 1/n_2^2)$ grafiğinin eğiminden Rydberg sabiti R_H tayin edilmektedir.

3.1.8. Davisson-Germer Deneyi

Clinton Davisson ve Lester Germer, 1927 yılında nikel kristalinden saçılan elektronların kırınım deseni oluşturduğunu keşfetmiştir. Bu bulgu, de Broglie'nin 1924'te önerdiği madde dalgaları hipotezinin doğrudan deneysel kanıtıdır (Davisson & Germer, 1927). Her iki araştırmacı da bu keşif nedeniyle 1937 Nobel Fizik Ödülü'ne layık görülmüştür.

Simülasyonda Ni kristalin kafes parametresi d = 0.215 nm sabit tutulurken hızlandırma gerilimi *V*, 10–300 V arasında değiştirilmektedir. Her *V* değeri için de Broglie dalga boyu ve Bragg kırınım açısı hesaplanmaktadır:

Tepe (pik) açısı ölçümünden hesaplanan Planck sabiti değeri, fotoelektrik etki deneyinden bağımsız olarak elde edilen ve CODATA değeriyle karşılaştırılabilen ikinci bir h tayinini temsil etmektedir.

3.1.9. Rutherford Saçılması Deneyi

Ernest Rutherford, 1909 yılında Geiger ve Marsden ile birlikte gerçekleştirdiği altın folyo deneyi sonucunda atomun yoğun ve küçük bir çekirdekten oluştuğunu keşfetmiştir. Bu keşif, Thomson atom modelinin yerini alan Rutherford atom modelinin temelini atmıştır (Rutherford, 1911). Alfa parçacıklarının geniş açılarda saçılması, pozitif yükün küçük bir hacimde yoğunlaşmış olduğunun kanıtıdır.

Simülasyon, Monte Carlo yöntemiyle rastgele etki parametreli (b) alfa parçacıklarını ateşlemekte ve her parçacık için Coulomb saçılmasını hesaplamaktadır. Hedef çekirdeğinin proton sayısı Z (1–82) ve alfa parçacığı enerjisi (0.5–10 MeV) ayarlanabilmektedir.

Saçılma açısı histogramının $N(\theta) - 1/\sin^4(\theta/2)$ grafiğiyle doğrusal uyumu, Rutherford dağılımının deneysel doğrulamasını simülasyon ortamında yeniden üretmektedir. Radyoaktif kaynak içermeyen bu yapı, deneyi radyasyon güvenliği kaygısı olmaksızın gerçekleştirilebilir kılmaktadır.

3.1.10. Stern-Gerlach Deneyi

Otto Stern ve Walther Gerlach, 1922 yılında inhomogen bir manyetik alan içinden gümüş atom demeti geçirerek uzay kuantizasyonunu ve spin açısal momentumunu deneysel olarak kanıtlamıştır (Gerlach & Stern, 1922). Bu deney, spin s = 1/2 kuantum sayısının ve Bohr magnetonunun doğrudan gösterimidir.

Simülasyonda manyetik alan gradyanı dB/dz (10–1000 T/m), mıknatıs uzunluğu ve atom hızı (sıcaklıkla Maxwell-Boltzmann dağılımına göre belirlenen) bağımsız olarak ayarlanabilmektedir. Gümüş atomlarının iki farklı band oluşturması; spin kuantizasyonunun görsel göstergesidir. $\Delta z - dB/dz$ grafiğinin eğiminden Bohr magnetonu μB tayin edilmektedir.

3.2. Deneylerin Karşılaştırmalı Değerlendirmesi

Tablo 3'te sistem içindeki on deney modülü; simüle edilen temel fizik ilkesi, tayin edilebilecek fizik sabiti, kullanılan analiz yöntemi ve beklenen ölçüm doğruluğu bakımından karşılaştırmalı olarak özetlenmektedir.

Tablo 3: Modern Fizik Sanal Laboratuvarı deney modüllerinin karşılaştırmalı özeti.

Deney	Temel Fizik İlkesi	Tayin Edilen Sabit	Analiz Yöntemi	Beklenen Hata
Fotoelektrik Etki	Foton-elektron etkileşimi	h (Planck sabiti)	Vs − f doğrusal regresyon	< %1
Karacisim Işıması	Planck kuantum dağılımı	b (Wien sabiti)	λmax − 1/T regresyon	< %0.5
Franck-Hertz	Atom enerji seviyeleri	E_0 uyarılma enerjisi	I−V minimum aralığı	< %2
Compton Saçılması	Foton-elektron çarpışması	λC (Compton d. boyu)	Δλ − (1−cosθ) regresyon	< %1
Millikan Yağ Damlası	Yük kuantizasyonu	e (elementsel yük)	GCD analizi	< %3
Çift Yarık	Dalga-parçacık dualitesi	λ (dalga boyu)	Saçak aralığı ölçümü	< %2
Hidrojen Spektrumu	Bohr atom modeli	RH (Rydberg sabiti)	$1/\lambda - (1/n_1^2 - 1/n_2^2)$ reg.	< %0.5
Davisson-Germer	Madde dalgaları (de Broglie)	h (Planck sabiti)	θpeak − V Bragg analizi	< %2
Rutherford Saçılması	Çekirdek modeli	Dağılım teyidi	N(θ) histogram fit	İstatistiksel
Stern-Gerlach	Spin kuantizasyonu	μB (Bohr magnetonu)	Δz − dB/dz regresyon	< %3

3.3. Sistemin Eğitimsel Yaklaşımı

Literatürde sanal fizik laboratuvarlarının etkinliği üzerine yapılan araştırmalar, birkaç temel eğitimsel ilkenin öne çıktığını göstermektedir. Wieman ve Perkins (2005), iyi tasarlanmış simülasyonların aktif öğrenmeyi desteklediğini, kavram yanılgılarını gidermede etkili olduğunu ve öğrencinin bilim insanı gibi düşünmesini teşvik ettiğini bildirmiştir. Perkins ve arkadaşları (2006) ise anında geri bildirim, çoklu temsil (animasyon + grafik + sayısal değer) ve öz keşif fırsatlarının kavramsal anlama üzerindeki olumlu etkisine dikkat çekmiştir.

Modern Fizik Sanal Laboratuvarı bu ilkeleri şu mekanizmalarla hayata geçirmektedir: (1) **Anında geri bildirim**: Parametre değişikliklerine karşılık gelen fiziksel büyüklükler (<16 ms gecikmeyle) LCD panellerde güncellenmektedir. (2) **Çoklu temsil**: Aynı fenomen animasyon (görsel), grafik (analitik) ve sayısal (LCD) düzeylerde eş zamanlı sunulmaktadır. (3) **Öz keşif**: Öğrenci, frekans artırıldığında durdurucu gerilimin değişip değişmediğini ya da ışık şiddetinin ne etkisi olduğunu bizzat keşfedebilmektedir. (4) **Doğrulama döngüsü**: Toplanan verilerden CODATA değerleriyle karşılaştırılabilir sabitler tayin edilmekte ve ölçüm kalitesi R^2 metriğiyle değerlendirilmektedir.

4. SONUÇLAR

Bu çalışmada; web tarayıcılarında herhangi bir kurulum gerektirmeksizin çalışan, CODATA 2018 sabitleriyle kalibre edilmiş ve on klasik modern fizik deneyini tek bir platformda toplayan interaktif bir sanal laboratuvar sistemi tasarlanmıştır. Sistem, aşağıdaki özgün katkıları sunmaktadır:

1. CODATA 2018 doğruluğunda fizik motoru: Tüm hesaplamalar merkezi sabit veri tabanı üzerinden yürütülmekte; Planck sabiti, Boltzmann sabiti ve elementsel yük gibi temel büyüklükler 2018 CODATA standart değerleriyle tam uyumlu olarak kullanılmaktadır.
2. On deneyi kapsayan entegre platform: Fotoelektrik etkiden Stern-Gerlach'a uzanan geniş kapsam, modern fiziğin tarihsel ve kavramsal bütünlüğünü tek bir öğrenme ortamında yansıtmaktadır.
3. Gerçek zamanlı regresyon analizi ve sabit tayini: Her deney için öğrenci, gerçek laboratuvar protokolüne benzer biçimde veri toplamakta; doğrusal regresyon yöntemiyle temel sabitleri bizzat tayin edebilmektedir.
4. Altı bölümlü konu anlatımı modülü: Tarih, teori, prosedür, sonuçlar, uygulamalar ve hata analizi sekmeleri, sistemi saf simülasyondan bütünleşik bir öğrenme ortamına dönüştürmektedir.
5. CSV veri dışa aktarımı: Toplanan veri setleri standart formatta indirilerek harici istatistiksel analizlerde (spreadsheet, Python, R) kullanılabilmektedir.

Sistemin gelecekteki geliştirme yönleri arasında; kullanıcı veri analizlerinin kalıcı olarak saklanması ve paylaşılabilmesi, öğrenci performans izleme modülü eklenmesi, çoklu kullanıcı iş birliği özelliği ve ek deney modülleri (fotoiyonizasyon, tünel olayı, kuantum parçacık kutusu) yer almaktadır. Önerilen sistemin fizik eğitimi araştırmacıları ve öğretim elemanlarının kullanımına sunulması ve etkinliğinin kontrol gruplu çalışmalarla sınanması, ilerleyen dönemdeki araştırma gündemimizi oluşturmaktadır.

KAYNAKLAR

Bohr, N. (1913). On the constitution of atoms and molecules. *Philosophical Magazine*, 26(151), 1–25. https://doi.org/10.1080/14786441308634993

Compton, A. H. (1923). A quantum theory of the scattering of x-rays by light elements. *Physical Review*, 21(5), 483–502. https://doi.org/10.1103/PhysRev.21.483

Davisson, C., & Germer, L. H. (1927). Diffraction of electrons by a crystal of nickel. *Physical Review*, 30(6), 705–740. https://doi.org/10.1103/PhysRev.30.705

de Broglie, L. (1924). Recherches sur la théorie des Quanta. *Annales de Physique*, 3, 22–128. https://doi.org/10.1051/anphys/192510030022

Einstein, A. (1905). Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik*, 17(6), 132–148. https://doi.org/10.1002/andp.19053220607

Feynman, R. (1965). The Feynman lectures on physics volome III : quantum mechanics / Richard P. Feyman, Robert B. Leighton, Marhew Sands. , *1965*, 1-99.

Franck, J., & Hertz, G. (1914). Über Zusammenstöße zwischen Elektronen und Molekülen des Quecksilberdampfes und die Ionisierungsspannung desselben. *Deutsche Physikalische Gesellschaft Verhandlungen*, 16, 457–467.

Gamage, K., Wijesuriya, D., Ekanayake, S., Rennie, A., Lambert, C., & Gunawardhana, N. (2020). Online Delivery of Teaching and Laboratory Practices: Continuity of University Programmes during COVID-19 Pandemic. *Education Sciences*. https://doi.org/10.3390/educsci10100291.

Gerlach, W., & Stern, O. (1922). Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld. *Zeitschrift für Physik*, 9(1), 349–352. https://doi.org/10.1007/BF01326983

Millikan, R. A. (1913). On the elementary electrical charge and the Avogadro constant. *Physical Review*, 2(2), 109–143. https://doi.org/10.1103/PhysRev.2.109

Millikan, R. A. (1916). A direct photoelectric determination of Planck's "h". *Physical Review*, 7(3), 355–388. https://doi.org/10.1103/PhysRev.7.355

Perkins, K., Adams, W., Dubson, M., Finkelstein, N., Reid, S., Wieman, C., & LeMaster, R. (2006). PhET: Interactive simulations for teaching and learning physics. *The Physics Teacher*, 44(1), 18–23. https://doi.org/10.1119/1.2150754

Planck, M. (1901). Ueber das Gesetz der Energieverteilung im Normalspectrum. *Annalen der Physik*, 309(3), 553–563. https://doi.org/10.1002/andp.19013090310

Radhamani, R., Kumar, D., Nizar, N., Achuthan, K., Nair, B., & Diwakar, S. (2021). What virtual laboratory usage tells us about laboratory skill education pre- and post-COVID-19: Focus on usage, behavior, intention and adoption. *Education and Information Technologies, 26*, 7477 - 7495. https://doi.org/10.1007/s10639-021-10583-3.

Rutherford, E. (1911). The scattering of α and β particles by matter and the structure of the atom. *Philosophical Magazine*, 21(125), 669–688. https://doi.org/10.1080/14786440508637080

Tiesinga, E., Mohr, P. J., Newell, D. B., & Taylor, B. N. (2021). CODATA recommended values of the fundamental physical constants: 2018. *Reviews of Modern Physics*, 93(2), 025010. https://doi.org/10.1103/RevModPhys.93.025010

Wieman, C., & Perkins, K. (2005). Transforming physics education. *Physics Today*, 58(11), 36–41. https://doi.org/10.1063/1.2155756

TurkicSign: Türk Dünyası İçin Uçtan Uca Tarayıcı Tabanlı Gerçek Zamanlı İşaret Dili Tanıma Sistemi

Murat Fazlı *Bilgi İşlem Dairesi Başkanlığı Yazılım Mühendisi Kırgızistan-Türkiye Manas Üniversitesi, Bişkek, Kırgızistan*
muratfazliev777@gmail.com

Özet-Türk dünyasının hızla dijitalleştiği günümüzde, işitme engelli bireyler için kapsayıcı bir ekosistem oluşturmak kritik bir ihtiyaçtır. Bu çalışma, "Dilde, Fikirde, İşte Dijital Birlik" vizyonu doğrultusunda geliştirilen ve işaret dilini gerçek zamanlı olarak metne ve sese çeviren "TurkicSign" yapay zeka platformunu sunmaktadır. Geleneksel bulut tabanlı sistemlerin aksine, bu projenin temel yeniliği, TensorFlow.js ve WebGL hızlandırması kullanılarak yapay zeka model çıkarım (inference) süreçlerinin doğrudan istemci tarafında (web tarayıcısında) gerçekleştirilmesidir. Sistem, MediaPipe çerçevesi kullanılarak el ve postür eklem noktalarının (landmarks) uzamsal-zamansal (spatio-temporal) analizini yapmakta ve dinamik jest sınıflandırması için Tekrarlayan Sinir Ağları (RNN) mimarisini kullanmaktadır. Mevcut ön prototip (Proof of Concept), temel iletişim için öncelikli olan 5 farklı jesti tanımakta ve Kırgız Türkçesi ile Türkiye Türkçesinde anlık çıktı vermektedir. Gerçek zamanlı doğrulama testlerinde modelin genel doğruluk oranı (accuracy) %85 olarak ölçülmüştür. Elde edilen bulgular, sistemin düşük gecikme süresi (low latency) ve yüksek veri gizliliği sağladığını kanıtlamaktadır. Bu teknik altyapı, Türk dilleri için geniş kapsamlı bir işaret dili veri setinin (korpus) oluşturulmasına ve gelecekte 3D Yapay Zeka Avatarları aracılığıyla çift yönlü iletişime zemin hazırlamaktadır.

Anahtar Kelimeler: Yapay Zeka, İşaret Dili Tanıma, TensorFlow.js, Uzamsal-Zamansal Analiz, MediaPipe, İnsan-Bilgisayar Etkileşimi.

I. GIRIŞ

Dijital dönüşüm çağında, işitme engelli bireylerin sosyal, akademik ve ekonomik hayata bağımsız bir şekilde entegre olabilmesi, erişilebilir yapay zeka (YZ) teknolojilerinin geliştirilmesine bağlıdır. Türk dünyasında, farklı lehçe ve şiveler konuşulmasına rağmen, işitme engelliler için ortak bir dijital iletişim platformunun (dijital köprü) eksikliği ciddi bir engel teşkil etmektedir.

Bilgisayarlı görü (computer vision) alanındaki son gelişmeler, işaret dili tanıma (Sign Language Recognition - SLR) sistemlerinde önemli sıçramalara yol açmıştır. Ancak mevcut sistemlerin çoğu yüksek işlem gücü gerektiren sunucu tabanlı (server-side) mimarilere dayanmaktadır. Bu durum, veri gizliliği ihlali risklerini artırmakta ve ağ gecikmeleri nedeniyle gerçek zamanlı akıcı iletişimi zorlaştırmaktadır.

Bu çalışmada sunulan "TurkicSign" platformu, bu sorunlara yenilikçi bir çözüm getirmektedir. Proje, donanımdan bağımsız çalışabilen, bulut maliyetlerini ortadan kaldıran ve tamamen web tarayıcısı üzerinden gerçek zamanlı çıkarım (client-side inference) yapabilen uçtan uca bir mimari önermektedir. Makalenin devamında, sistemin uzamsal-zamansal özellik çıkarma metodolojisi, sinir ağı mimarisi ve gerçek zamanlı performans analizi detaylandırılmaktadır.

II. SISTEM MIMARISI VE METODOLOJI

A. Uzamsal-Zamansal Özellik Çıkarımı (Spatio-Temporal Feature Extraction)

İşaret dili, statik bir görüntüden ziyade zaman içinde değişen dinamik hareketler dizisidir. Bu nedenle, ham RGB video kareleri yerine iskelet tabanlı (skeleton-based) yaklaşım benimsenmiştir. Sistemde, Google MediaPipe Holistic boru hattı (pipeline) kullanılmıştır.

- Her bir video karesinden (frame). Sağ ve sol el için 21'er adet 3B (x, y, z) eklem noktası (hand landmarks),
- Üst vücut postürü için 33 adet 3B referans noktası (pose landmarks) çıkarılmaktadır.

Elde edilen uzamsal koordinatlar, her bir kare için tek boyutlu bir özellik vektörüne (feature vector) düzleştirilir (flattening). Belirli bir zaman aralığındaki (örneğin $T=30$ kare) vektörler birleştirilerek, modelin girdisi olan uzamsal-zamansal matrisler oluşturulur. Bu yöntem, aydınlatma değişiklikleri ve arka plan gürültüsü (background noise) gibi dış faktörlerin sınıflandırma üzerindeki olumsuz etkisini minimize eder.

B. Yapay Sinir Ağı Mimarisi

Ardışık (zaman serisi) verilerin analizi için Standart Evrişimli Sinir Ağları (CNN) yerine Tekrarlayan Sinir Ağları (RNN) tercih edilmiştir. Uzun süreli bağımlılıkları (long-term dependencies) öğrenebilmek ve kaybolan gradyan (vanishing gradient) problemini aşmak amacıyla model, Uzun Kısa Süreli Bellek (LSTM) katmanları üzerine inşa edilmiştir.
Modelin mimarisi şu şekildedir:

1. **Girdi Katmanı:** Ardışık iskelet koordinat matrislerini alır.
2. **Derin LSTM Katmanları:** Zaman içindeki hareket dinamiklerini (ellerin yönü, hızı ve vücutla olan etkileşimi) analiz eder.
3. **Tam Bağlantılı Katmanlar (Dense Layers):** ReLU aktivasyon fonksiyonu kullanılarak doğrusal olmayan ilişkiler modellenir.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

4. **Çıktı Katmanı:** Softmax aktivasyon fonksiyonu ile sınıflara ait olasılık dağılımlarını hesaplar.

Model öncelikle Python ortamında (TensorFlow/Keras) action.h5 formatında eğitilmiş ve ağırlıkları optimize edilmiştir.

C. Tarayıcı Tabanlı Uç Bilişim (Edge Computing in Browser)

Projenin en kritik teknik yeniliği, eğitilmiş modelin ağ topolojisinin ve ağırlıklarının (model.json ve group1-shard1of1.bin) TensorFlow.js (TFJS) formatına dönüştürülmesidir. Çıkarım işlemi, kullanıcının cihazındaki grafik işlem birimini (GPU) WebGL API'si üzerinden kullanarak doğrudan tarayıcı içerisinde (JavaScript ile) gerçekleştirilir. Bu "sıfır sunucu" (zero-server) yaklaşımı, bant genişliği tüketimini düşürür, %100 kullanıcı gizliliği sağlar ve milisaniyeler seviyesinde (low-latency) anlık tepki süresi sunar.

III. Deneysel Bulgular ve Performans Analizi

Sistemin uygulanabilirliğini test etmek amacıyla, 5 kritik jestten (Kardeş, Kırgızistan, Teşekkürler/Rahmat, Selam, Türkiye) oluşan bir pilot veri seti hazırlanmıştır. Modelin performansı, laboratuvar ortamı dışında, farklı aydınlatma koşullarında ve farklı donanımlara sahip standart dizüstü bilgisayarlarda test edilmiştir.

Gerçek zamanlı test sonuçlarına göre modelin genel doğruluk oranı (Accuracy) %85 olarak ölçülmüştür.

TABLE I. Jest Sınıflarına Göre Sınıflandırma Performansı

Jest Sınıfı	Hassasiyet (Precision)	Duyarlılık (Recall)	F1-Skoru
Selam	0.88	0.89	0.88
Rahmat (Teşekkürler)	0.86	0.85	0.85
Kırgızistan	0.85	0.86	0.85
Kardeş	0.84	0.82	0.83
Türkiye	0.82	0.83	0.82
Ağırlıklı Ortalama	0.85	0.85	0.85

Karmaşıklık matrisi (Confusion Matrix) analizi, özellikle "Kardeş" ve "Türkiye" jestlerinde el pozisyonlarının başlangıç noktalarındaki benzerlikler nedeniyle minör yanlış sınıflandırmaların yaşandığını, ancak genel akıcılığın korunduğunu göstermiştir. Tanınan her jest, anlık olarak web arayüzünde hem metne dökülmekte hem de sesli (audio playback) olarak ifade edilmektedir.

IV. Tartışma ve Gelecek Çalışmalar (Vizyon 2026-2027)

Elde edilen %85'lik doğruluk oranı, web tabanlı bir uç bilişim (edge AI) sistemi için oldukça umut vericidir. Projenin sonraki fazlarında aşağıdaki teknik geliştirmeler hedeflenmektedir:

- **Genişletilmiş Türk Dünyası Korpusu:** Yalnızca Kırgızca ve Türkçe değil, Kazakça, Özbekçe ve Azerbaycan Türkçesi işaret dillerini de kapsayan, on binlerce video örneğinden oluşan açık kaynaklı bir veri setinin oluşturulması.
- **Transformer Mimarilerine Geçiş:** Hareket dizilerinin daha bağlamsal analizi için LSTM yerine Vision Transformers (ViT) veya Spatial-Temporal Graph Convolutional Networks (ST-GCN) kullanılması.
- **Çift Yönlü İletişim ve AI Avatar:** Mevcut sistem (İşaretten $\rightarrow$ Sese/Yazıya) tek yönlüdür. Gelecekte, işiten bireylerin konuşmalarını anlık olarak işaret diline çeviren ve bunu ekranda 3 Boyutlu Foto-Gerçekçi (Photo-realistic) Yapay Zeka Avatarları ile gösteren "Tam Zamanlı Dijital Tercüman" entegrasyonu sağlanacaktır.

V. Sonuç

TurkicSign, engelsiz bir dijital Türk dünyası inşa etme hedefinde atılmış stratejik bir teknik adımdır. İskelet tabanlı özellik çıkarımı ve tarayıcı içi makine öğrenimi (TFJS) yaklaşımlarını birleştiren bu çalışma; yüksek sunucu maliyetleri, gecikme ve veri gizliliği gibi geleneksel işaret dili tanıma sistemlerinin temel zayıflıklarını başarıyla gidermiştir. Geliştirilen prototipin başarılı sonuçları, yapay zekanın sadece endüstriyel bir araç olmadığını, aynı zamanda toplumsal entegrasyonu ve kültürel/dilsel bütünlüğü sağlayan en güçlü köprülerden biri olduğunu kanıtlamaktadır.

Referanslar

[1] C. Lugaresi et al., "MediaPipe: A Framework for Building Perception Pipelines," arXiv preprint arXiv:1906.08172, 2019.

[2] D. Smilkov, N. Thorat, B. Assis, et al., "TensorFlow.js: Machine Learning for the Web and Beyond," Proceedings of the 2nd SysML Conference, Palo Alto, CA, USA, 2019.

[3] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[4] J. Qin, et al., "Fusing Wearable IMUs and Smartphone Camera for Sign Language Recognition," IEEE Transactions on Human-Machine Systems, 2021.

Re-examining 12-Lead ECG Benchmarks: Anti-Aliased Decimation, Label Granularity, and Cross-Dataset Transfer

Elaman Nazarkulov
Department of Computer Engineering
Kyrgyz–Turkish Manas University, Bishkek, Kyrgyzstan
elaman.job@gmail.com

***Abstract*—Anti-aliased decimation of the 12-lead electrocardiogram (ECG) input from 5,000 to 500 samples raises a plain residual 1D convolutional neural network from 88.43% to 97.34% test accuracy on Chapman–Shaoxing — exceeding the 94.8% accuracy reported for an attention-hybrid successor without any change to the model, the loss, or the augmentation recipe. We use this controlled in-distribution baseline to interrogate cross-dataset transfer to PTB-XL and find three results that the standard 5-class superclass collapse obscures: (i) re-evaluating the same trained representation at an 11-label intersection of SNOMED CT and SCP-ECG codes lifts source-only macro area under the ROC curve (AUC) from 0.556 to 0.838; (ii) per-class decision-threshold calibration on PTB-XL fold 9 raises accuracy from 0.584 to 0.664 at zero training cost; (iii) five epochs of PTB-XL fine-tuning at the intersection granularity produce macro-AUC 0.948 on held-out fold 10 — exceeding the 0.925 macro-AUC obtained by training directly on PTB-XL in the published benchmark. We argue that input length and evaluation-time label granularity are under-reported design variables in ECG benchmarks, and that aspirational numbers in the literature may partly reflect their omission rather than the architectural contributions claimed.**

***Index Terms*—electrocardiogram, 12-lead ECG, deep learning, 1D-CNN, anti-aliased decimation, cross-dataset transfer, PTB-XL, label granularity, threshold calibration, fine-tuning, multi-class classification.**

I. INTRODUCTION

Deep convolutional networks now reach cardiologist-level performance on 12-lead ECG interpretation [1], [2], [3]. Nearly every published pipeline consumes the signal at its acquisition rate, most commonly 500 Hz producing 5,000 samples per lead for a 10 s segment. The choice is treated as given: data augmentation [6], [7], support-node interpolation [8], [9], and hybrid recurrent architectures [10] are evaluated on top of this fixed input.

A baseline 1D-CNN trained on the Chapman–Shaoxing corpus [5] in a prior phase of this work reached only 88.43% test accuracy with macro F1 0.8713 and eleven labels collapsing below F1 < 0.60. The natural reaction is to add attention, recurrent layers, and a tuned focal loss [11]; the literature indeed reports a 94.8%-accuracy CNN–LSTM hybrid [10] as the next step. We test a contrary hypothesis: that the 5,000-sample input already carries more temporal redundancy than the CNN can usefully exploit, and that simple anti-aliased decimation to 500 samples preserves every diagnostically relevant feature while concentrating gradient signal on them.

Beyond this in-distribution result, a second question arises: does the trained model transfer to a different 12-lead corpus? Source-only evaluation on PTB-XL [4] at the conventional 5-class superclass level gives macro-AUC 0.556, leading to a natural conclusion that the model has overfit to Chapman–Shaoxing. We show that this conclusion is partly an artefact of the evaluation: the same representations, scored at a finer 11-label SNOMED CT ↔ SCP-ECG intersection, already produce macro-AUC 0.838; one threshold per class fitted on PTB-XL fold 9 raises accuracy from 0.584 to 0.664 at zero training cost; and five epochs of target-domain fine-tuning push macro-AUC to 0.948 on held-out fold 10, exceeding Strodthoff et al. [3] who trained directly on PTB-XL.

Contributions. This paper contributes:

1. A controlled comparison of input lengths {5,000, 1,000, 500} on Chapman–Shaoxing with identical model, augmentation, optimiser, seed, and data split, isolating the 8.91 pp accuracy improvement attributable solely to anti-aliased decimation.
2. Cross-dataset evaluation of the resulting model on PTB-XL at three label-granularities (78-class collapsed, 5-class native, 11-class intersection), demonstrating that source-only macro-AUC depends strongly on evaluation-time label resolution.
3. Per-class decision-threshold calibration as a no-retraining adaptation lever that closes the source-only argmax gap by 8 pp.
4. A target-domain fine-tuning result that exceeds the PTB-XL native baseline of [3] at the 11-label intersection.

II. RELATED WORK

Deep ECG classification. Rajpurkar et al. [1] and Hannun et al. [2] trained 34-layer CNNs on 91,232 ECGs and matched cardiologist-level rhythm recognition. Strodthoff et al. [3] benchmark CNN and RNN models on PTB-XL using a downsampled 100 Hz (1,000-sample) input for compute reasons and reach macro-AUC 0.925. Oh et al. [10] report

94.8% accuracy on variable-length heartbeats with a CNN–LSTM hybrid.

Augmentation. Iwana and Uchida [6] survey time-series augmentation. GAN-based synthesis [7] and support-guided interpolation of fiducial points [8], [9] are dominant ECG-specific recipes.

Imbalance. Focal loss [11] and inverse-frequency re-weighting are the standard responses to severe class imbalance. Class-balanced loss based on effective number of samples [12] extends this to long-tailed targets.

Input length. In every cited study, input length and sample rate are configured once and not revisited. To our knowledge no prior large-scale 12-lead ECG work reports a controlled input-length ablation as its primary result.

Cross-dataset transfer. Strodthoff et al. [3] provide the canonical PTB-XL benchmark but do not address transfer from independently-sourced training corpora. Per-class threshold adjustment is a known technique in calibrated multi-class classification but has not, to our knowledge, been applied to cross-dataset ECG benchmarks.

III. METHOD

A. Datasets and preprocessing

We use Chapman–Shaoxing [5] (45,152 records at 500 Hz, 10 s each, 78 SNOMED-mapped diagnostic labels) for training and in-distribution test; PTB-XL [4] (21,388 records with at least one diagnostic SCP code; native distribution NORM ≈ 42%, MI ≈ 26%, STTC ≈ 15%, CD ≈ 14%, HYP ≈ 2.5%) for cross-dataset evaluation, using its canonical stratified split (folds 1–8 train, fold 9 validation, fold 10 held-out test) following [3].

Preprocessing is identical across both corpora and across all configurations: (1) resample to 500 Hz (sinc interpolation); (2) bandpass 0.5–150 Hz (Butterworth order 4) and 50 Hz notch for power-line interference; (3) high-pass 0.5 Hz for baseline-wander removal; (4) per-lead Z-score normalisation with ±3σ clipping; (5) fixed 10 s segmentation to [12 × 5,000]; (6) signal-quality index ≥ 0.85 (42,513 of 45,152 Chapman records retained); (7) anti-aliased decimation (new) — applied only in the non-baseline configurations; (8) support-node augmentation [9]: 3× for common classes, 10× for rare classes, target 4,500 samples per class. A stratified 68/12/20 train/val/test split is fixed by a single seed across all in-distribution configurations.

B. Anti-aliased decimation

Let x ∈ ℝ^(12×N) with N = 5,000. The decimation step is a single call to scipy.signal.decimate [14]:

$$x' = \text{decimate}(x, q, \text{ftype} = \text{iir}, n = 8, \text{zero_phase} = \text{True}) \quad (1)$$

with q ∈ {1, 5, 10} corresponding to N′ ∈ {5,000, 1,000, 500} output samples. The IIR low-pass is a Chebyshev-I filter of order eight applied in forward–backward mode to preserve phase; its cutoff sits at the new Nyquist frequency, removing any spectral content that would otherwise alias into the QRS band [13]. No other pipeline step changes between configurations.

C. Model

A residual 1D-CNN: five convolutional blocks with filter counts [64, 128, 256, 512, 512], kernel sizes [16, 16, 16, 8, 8], batch normalisation, ReLU, and max-pooling; global average pooling; two dense layers (256 → K) with dropout 0.5, where K is the number of target classes (K = 78 in-distribution, K = 11 for the intersection). Total: 3.72 M parameters. Loss: focal loss [11] with $\gamma = 2$ and class weights, combined with label-smoothed cross-entropy ($\varepsilon = 0.1$). Optimiser: Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$, learning rate 10^{-3}, ReduceLROnPlateau patience 5). Batch size 64. Early stopping on validation loss with patience 10 and max 100 epochs.

D. Cross-dataset adaptation

For cross-dataset evaluation we use three adaptation regimes on the trained model:

- **Source-only:** forward-pass the model on PTB-XL with no adaptation. Reported under three label-granularities (78 fine collapsed, 5 native superclass, 11 SNOMED-SCP intersection).
- **Per-class threshold calibration:** for each class c, the threshold $\tau_c \in [0.02, 0.99]$ is chosen on PTB-XL fold 9 to maximise one-vs-rest F1. At test time, predictions become argmax_c ($p_c - \tau_c$) with a fallback to plain argmax_c p_c when every margin is negative. No retraining; eleven scalar hyperparameters.
- **Target-domain fine-tuning:** five epochs of PTB-XL train folds 1–8 at learning rate 10^{-5} with AdamW [15], label smoothing 0.1, batch 64, no frozen layers. Best validation macro-F1 on fold 9 is kept.

E. Hardware

A single NVIDIA RTX 5090 (34.19 GiB, CUDA 12.8) with mixed-precision (FP16) training.

IV. IN-DISTRIBUTION RESULTS

A. Headline comparison

Table I reports test accuracy, macro F1, single-sample inference latency, and mean softmax confidence across four input-length configurations. All other variables are fixed.

TABLE I

Chapman–Shaoxing test-set metrics across four input-length configurations.

Configuration	Acc.	Macro F1	Inference	Conf.
len = 5,000 (baseline)	0.8843	0.8713	89.88 ms	0.1289
len = 1,000	0.9722	0.9716	26.14 ms	0.6888
len = 500	0.9734	0.9737	27.20 ms	0.7623
len = 500, 4	**0.9738**	**0.9744**	43.50 ms	0.6959

workers

Reducing the input length from 5,000 to 500 samples improves test accuracy by 8.91 pp and macro F1 by 10.24 pp, with 3.3× faster inference. The second decimation step (1,000 → 500) contributes only 0.12 pp accuracy, indicating that the bulk of the effect is captured at 1,000 samples.

B. Per-class recovery

The eleven labels with F1 < 0.60 in the len = 5,000 baseline all recover to F1 ≥ 0.95 at len = 500 (Table II; LVH moves from 0.022 to ≥ 0.99). The recovery is uniform across diverse arrhythmia families (atrial flutter, AV block, premature beats, hypertrophy), suggesting the effect is not class-specific.

TABLE II

Per-class F1 recovery from len = 5,000 to len = 500, top 6 entries.

Class	len = 5,000	len = 500	ΔF1
LVH	0.022	≥ 0.99	+0.97
Q wave abnormal	0.180	≥ 0.99	+0.81
Intraventricular block	0.286	≥ 0.98	+0.70
Atrioventricular block	0.324	0.984	+0.66
Premature atrial contraction	0.329	≥ 0.97	+0.64
Atrial fibrillation	0.436	≥ 0.95	+0.51

V. CROSS-DATASET VALIDATION ON PTB-XL

A. Three label-granularities

The in-distribution result of Section IV could mask overfitting to Chapman–Shaoxing. We evaluate the same trained model on PTB-XL fold 10 at three label-granularities, in order of decreasing coarseness:

- **78 → 5 collapse:** the 78-class trained labels are mapped at evaluation time to the five PTB-XL superclasses {NORM, MI, STTC, CD, HYP} [3].
- **Native 5-class:** a new model is trained from scratch on Chapman–Shaoxing with labels pre-collapsed to the same five superclasses. In-distribution accuracy drops from 0.973 to 0.908 as the task becomes coarser.
- **11-label intersection:** we hand-construct an 11-label intersection of SNOMED CT and SCP-ECG codes for which both datasets carry an unambiguous one-to-one mapping: SR, SBRAD, STACH, SARRH, AFIB, AFLT, 1AVB, RBBB, LBBB, LVH, PAC. Six clinically important codes from the original draft (RVH, 2AVB, 3AVB, ILBBB, WPW, PVC) are dropped because the Chapman subset contains fewer than ten records of each.

B. Source-only and adapted results

Table III summarises all configurations on PTB-XL fold 10.

TABLE III

PTB-XL fold 10 test metrics across seven cross-dataset configurations. Strodthoff 2020 [3]'s PTB-XL-native macro-AUC of 0.925 is the published reference; configuration 7 exceeds it at the 11-label granularity.

#	Configuration	n	Acc.	Macro F1	Macro AUC
1	78-class, 1-lead, super5-collapsed (source-only)	2,158	0.359	0.205	0.587
2	78-class, 12-lead, super5-collapsed (source-only)	2,158	0.366	0.205	0.599
3	5-class, 12-lead, super5-trained (source-only)	2,158	0.385	0.221	0.556
4	5-class, 12-lead, super5 + PTB-XL fine-tune	2,158	0.664	0.477	0.856
5	**11-class intersection (source-only)**	2,085	0.584	0.283	**0.838**
6	**11-class intersection + threshold calibration**	2,085	**0.664**	0.360	0.838
7	**11-class intersection + PTB-XL fine-tune**	2,085	**0.814**	**0.537**	**0.948**
—	Strodthoff 2020 [3] (PTB-XL-native, 71 SCP codes)	—	—	—	0.925

The cross-dataset gap is partly an artefact of label granularity at evaluation. Row 5 versus Row 3: holding the model representation constant (same architecture, same Chapman-only training, same 12 leads), changing the evaluation granularity from a 5-class super-class collapse to the 11-label intersection lifts source-only macro-AUC from 0.556 to 0.838 (+28.2 pp). The model's per-class probability rankings were never broken; the superclass collapse was averaging them into noise.

Per-class calibration alone closes the source-only argmax gap. Row 6 versus Row 5: the intersection model's ranking quality (AUC 0.838) does not translate into argmax accuracy 0.584 because the per-class decision thresholds inherited from Chapman are mis-aligned with PTB-XL's class prior. Fitting eleven scalar thresholds on PTB-XL fold 9 raises accuracy to 0.664 without changing AUC. The AFIB F1 alone jumps from 0.015 to 0.744 because the AFIB probabilities

were always sensible; only the argmax tiebreaker was losing them to AFLT. Per-class threshold calibration is the cheapest cross-dataset adaptation lever — no training, no new weights, eleven scalar hyperparameters fitted on the target-domain validation fold.

At fine granularity, target-domain fine-tuning exceeds the PTB-XL-native baseline. Row 7: five epochs of fine-tuning at LR 10^{-5} on PTB-XL train folds 1–8 produce macro-AUC 0.948 on fold 10 — exceeding Strodthoff et al. [3]'s 0.925 macro-AUC, obtained by training directly on PTB-XL with their full 71-SCP multi-label objective. The comparison is not fully apples-to-apples (we work at 11 superclass-like buckets, the cited benchmark at 71 fine SCP codes), but the direction is unambiguous: the same source representation, briefly adapted at the right label granularity, beats a from-scratch PTB-XL model at the 11-bucket task. Per-class F1 reaches ≥ 0.85 for SR, AFIB, LBBB, and RBBB after fine-tuning.

VI. DISCUSSION

A. Why 88% → 97%

Three forces compound to produce the observed in-distribution effect.

Receptive field. The last convolutional layer of our network has an effective receptive field of ≈ 2,048 input samples. At 5,000 samples this covers only ~40% of the window, yet the discriminating pattern — QRS morphology — fits in ≈ 100 samples at 500 Hz. After decimation to 500 samples, the same QRS fits in ≈ 10 samples and the receptive field spans the whole window, so local features and rhythm context are now both learnable.

Parameter economy. Network capacity (3.72 M parameters) is constant; ten-fold shorter input means those parameters are not spent on modelling redundant low-frequency variation between QRS complexes.

Anti-aliasing is load-bearing. A naïve strided-by-10 pooling produces a folded spectrum where QRS energy aliases into the low-frequency band, degrading rather than improving accuracy in preliminary tests. The Chebyshev anti-aliasing filter [13] is the difference between '+10 pp F1' and 'worse than baseline'.

B. What the in-distribution result does not show

The decimation effect does not imply that attention, recurrent layers, or focal loss are useless — it implies that they were measured against a baseline under-trained in the input dimension, so their reported contribution is an upper bound relative to a lower starting point. Revised contribution estimates against a decimate-500 baseline are part of our future work.

C. What the cross-dataset result adds

The Section V results reframe a putative cross-dataset failure as the joint outcome of two understated levers — evaluation label granularity and per-class threshold calibration — both of which operate at inference time on a fixed source-only representation. Five epochs of target-domain fine-tuning, at the right label granularity, then suffice to exceed the published PTB-XL-native macro-AUC at the intersection task. The in-distribution finding of Section IV (input length) and the cross-dataset findings of Section V (label granularity, calibration, target-domain data) are complementary; they bound where architectural levers can and cannot move accuracy on 12-lead ECG classification.

VII. LIMITATIONS AND FUTURE WORK

We have not characterised the decimation factor below 500 samples ($q > 10$), or the interaction between input length and deeper or attention-augmented models. The geometric-invariance argument may not hold for sub-diagnoses relying on high-frequency content (late potentials, micro-alternans) that are removed by the 25 Hz anti-aliasing cutoff at $q = 10$. The cross-dataset HYP-class collapse under naïve fine-tuning indicates that minimal target-domain exposure does not solve target-domain class imbalance; class-balanced loss [12] is the planned fix. Every reported result is single-seed; a multi-seed robustness study with reported variance bands is planned for the journal extension.

Planned follow-ups in order of expected payoff: (i) HYP-class recovery via class-balanced loss [12] or oversampling; (ii) combined-dataset multi-label training (Chapman–Shaoxing ∪ PTB-XL with sigmoid + BCE, targeting the 0.925 macro-AUC of [3] in a single model); (iii) label-taxonomy clean-up (78 → ~55 classes) and re-run; (iv) Attention-CNN-LSTM on decimate-500 input, evaluated against both source-only and fine-tuned baselines of Section V; (v) focal loss [11] with $\gamma \in \{1, 2, 3\}$; (vi) GradCAM and SHAP attribution on decimated input; (vii) edge deployment via INT8 quantisation on a Raspberry Pi 4 (< 100 ms/sample at < 1 pp accuracy loss).

VIII. CONCLUSION

Anti-aliased decimation of the 12-lead ECG input from 5,000 to 500 samples turns a plain 1D-CNN baseline into a model that exceeds the 94.8% accuracy target published for its attention-hybrid successor — without any change to the model, the loss, or the augmentation recipe. The single biggest lever available on the in-distribution Chapman–Shaoxing task was not the architecture but the input representation. The same trained model, re-evaluated at an 11-label intersection of SNOMED CT and SCP-ECG codes, transfers to PTB-XL fold 10 at macro-AUC 0.838 without any adaptation; per-class threshold calibration on fold 9 raises accuracy from 0.584 to 0.664 at zero training cost; and five epochs of fine-tuning push macro-AUC to 0.948, exceeding the PTB-XL-native baseline of Strodthoff et al. [3]. We argue that input length and evaluation-time label granularity are under-reported design variables in 12-lead ECG benchmarks, and that aspirational numbers in recent literature may partly reflect their omission rather than the architectural contributions claimed.

. REFERENCES

[1] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," arXiv:1707.01836, 2017.

[2] A. Y. Hannun et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," Nature Medicine, vol. 25, no. 1, pp. 65–69, 2019.

[3] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ECG analysis: benchmarks and insights from PTB-XL," IEEE J. Biomed. Health Inform., vol. 25, no. 5, pp. 1519–1528, 2020.

[4] P. Wagner et al., "PTB-XL, a large publicly available electrocardiography dataset," Scientific Data, vol. 7, no. 1, p. 154, 2020.

[5] J. Zheng et al., "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," Scientific Data, vol. 7, no. 1, p. 48, 2020.

[6] B. K. Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," PLOS ONE, vol. 16, no. 7, p. e0254841, 2021.

[7] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks," in Proc. IJCNN, 2017, pp. 1578–1585.

[8] X. Chen, Z. Wang, and M. J. McKeown, "Adaptive support-guided deep learning for physiological signal analysis," IEEE Trans. Biomed. Eng., vol. 68, no. 5, pp. 1573–1584, 2021.

[9] S. S. Xu, M.-W. Mak, and C. C. Cheung, "Support-guided augmentation for electrocardiogram signal classification," Biomed. Signal Process. Control, vol. 71, p. 103213, 2022.

[10] S. L. Oh, E. Y. Ng, R. S. Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," Comput. Biol. Med., vol. 102, pp. 278–287, 2018.

[11] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proc. ICCV, 2017, pp. 2980–2988.

[12] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in Proc. CVPR, 2019, pp. 9268–9277.

[13] A. V. Oppenheim and R. W. Schafer, Discrete-Time Signal Processing, 3rd ed. Pearson, 2009.

[14] P. Virtanen et al., "SciPy 1.0: fundamental algorithms for scientific computing in Python," Nature Methods, vol. 17, no. 3, pp. 261–272, 2020.

[15] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in Proc. ICLR, 2019.

A hybrid post-processing method for Cyrillic OCR output using Large Language Models

Zholdoshbek Tashiev
Kyrgyz-Turkish Manas University
Bishkek, Kyrgyz Republic
2552y0501001@manas.edu.kg

Rita Ismailova
Kyrgyz-Turkish Manas University
Bishkek, Kyrgyz Republic
rita.ismailova@manas.edu.kg

Digitizing texts in low-resource Cyrillic languages, such as Kyrgyz, Kazakh, and Uzbek, remains a challenge for current OCR systems. A key issue is that some letters, such as ө, ү, and ң, resemble standard Cyrillic letters but are rarely included in the training data. As a result, OCR models frequently confuse them, leading to a high error rate in practice. In this work, we explored a straightforward idea: instead of relying solely on OCR, we added a second step using a large language model to correct the output. This created a hybrid process—OCR first, followed by correction. We tested four OCR engines (Tesseract, EasyOCR, RapidOCR, and a Vision-based LLM) on several manually transcribed pages from a Kyrgyz book, which served as our ground truth. For the correction phase, we utilized the Qwen 35B-A3B model. The prompt was tailored for Kyrgyz and low-resource Cyrillic text, although this stage involved some trial and error. One unexpected finding was how sensitive the model was to input length. When we provided longer text segments, it sometimes produced odd corrections or even added content that wasn't present. However, when we divided the text into shorter, sentence-level segments, the outcomes improved significantly. In terms of results, the best configuration was Tesseract combined with Qwen. This setup lowered the Character Error Rate (CER) from 6.95% to 3.30% and the Word Error Rate (WER) from 25.76% to 12.85%. Interestingly, the Vision LLM alone, without a separate OCR step, achieved an even lower CER of 1.61%. We did not anticipate this, but it appears to perform well in this instance. Overall, it seems that relatively small adjustments, such as breaking the text into shorter segments and refining the prompt, can lead to meaningful improvements. This field of low-resource Cyrillic OCR remains underexplored, indicating that there is much more to investigate.

Keywords— OCR, post-processing, large language models, LLM, Kyrgyz language, CER, WER, text segmentation, low-resource languages.Introduction (Heading 1)

I. INTRODUCTION

Imagine you have a stack of old Kyrgyz books and want to digitize them. You photograph the pages, run a recognition program — and get text full of errors. The letter "ө" becomes "о", "ү" becomes "у", words merge or break apart. This is exactly the problem we encountered at the outset. OCR programs were designed primarily for English and Western European languages. Kyrgyz occupies a negligible share of their training data. As a result, even good OCR engines make far more errors on Kyrgyz texts than is acceptable [1, 5].

Preliminary experiments on recognizing Kyrgyz books revealed three specific problems that make raw OCR output unsuitable for further automated processing:

- **Problem 1 — High OCR error rate on Cyrillic texts of low-resource languages.** OCR regularly confuses letters ө, ү, ң with visually similar symbols о, у, н — simply because Cyrillic training data is nearly absent from most OCR systems [5]. The output is text where every third or fourth word is misspelled.
- **Problem 2 — Lack of specialized correctors for Cyrillic texts.** We reviewed dozens of papers and found none addressing Cyrillic low-resource languages — Kyrgyz, Kazakh, Uzbek. Moreover, Kyrgyz and Kazakh are agglutinative languages, where a single root can generate hundreds of word forms. No tools exist [4, 9, 10].
- **Problem 3 — LLM quality degradation on long text segments.** LLMs can understand context and correct errors — which sounds ideal. But when we tried feeding a full page, the model began "hallucinating": inventing words absent from the original. The longer the text, the worse the results. Kanerva et al. [4] call this the "segmentation effect."

We developed a method that addresses all three problems simultaneously. The following sections describe: what others have done (Section 2), our proposed approach (Section 3), and practical results (Sections 4–5).

II. LITERATURE REVIEW

Addressing Problem 1: Poor OCR Quality

The idea of correcting OCR errors with language models emerged several years ago. M. Hajiali [1] in 2023 was among the first to try GPT-4 for this purpose — on English medical texts. The result was convincing: WER dropped from 38.5% to 24.67%. A standard BERT model performed worse — 31.44%. The conclusion is straightforward: large language models understand context rather than simply consulting a dictionary, and therefore correct more effectively.

Subsequent work followed the same direction. Koynov et al. [9] in 2025 ran eight different LLMs on English newspapers — the best result came from GPT-4o mini: WER dropped from 18.63% to 4.40%. With fine-tuning on relevant data, WER falls to 3.22%.

Narbona and Ros (2025) [10] achieved the most striking result on low-quality Spanish documents: CER dropped from 27.07% to

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

4.32% (−84.0%) using Gemini 2.5 Pro. This demonstrates that LLMs can radically improve OCR output even when baseline error rates are high.

Summary for Problem 1: well solved for English and Spanish. However, none of these works addressed Cyrillic low-resource languages. The closest analogue is Finnish: there GPT-4o yielded −33.5% on WER, compared to −59.1% for English [4]. The situation for Kyrgyz is worse still, since Finnish at least appears frequently online — Kyrgyz does not.Units

Addressing Problem 2: Lack of Tools for Low-Resource Languages

Kanerva et al. [4] honestly tested whether LLM post-processing works for languages with limited internet presence. They chose Finnish. The paper's title says it all: "No Free Lunches." GPT-4o helped somewhat, while open models like Llama and Gemma were practically ineffective.

Alekseev and Turatali [5] in 2024 surveyed all NLP tools for Kyrgyz. The picture is bleak: Kyrgyz text is nearly absent from LLM training data, the specific characters ө, ү, ң are tokenized incorrectly, and no parallel corpora exist. The authors conclude that applying any LLM to Kyrgyz "as-is" yields unpredictable results.

Fritzen and Balreira (2025) [7] studied OCR post-processing for Portuguese and demonstrated the critical role of prompt engineering: clear instructions in the prompt substantially reduce the risk of text loss and introduction of new errors. Araújo et al. (2025) [11] tested over ten LLMs on Portuguese manuscripts and found that the specialized Portuguese model Sabia 3 underperformed mid-sized multilingual models.

Summary: Problem 2 is partially solved for European languages. For Cyrillic low-resource languages — no research has been conducted at all. Existing approaches show that careful prompt engineering and the choice of multilingual LLMs matter more than model size.

Addressing Problem 3: LLM Degradation on Long Segments

This is perhaps the best-studied of the three. Kanerva et al. [4] showed clearly: the shorter the text segment, the better the model corrects it. Approximately 50–100 words is optimal. Beyond that, the model becomes confused and introduces new errors instead of fixing existing ones.

Shi et al. [2] devised a clever approach: rather than running the entire text through the LLM, only segments where the OCR itself is uncertain are processed. This improved accuracy by 17%.

Koynov et al. [9] encountered an especially unpleasant case: on complex, noisy German text, Llama-3.3 not only failed to help — it made things worse. CER rose from 25% to 36%. The model simply got lost in the stream of errors.

Summary across all three problems: the literature converges on one principle — break text into short segments. The specific threshold for Cyrillic low-resource languages had not been investigated — we were the first.

TABLE I. OVERVIEW OF KEY RESULTS AND UNRESOLVED ISSUES IN THE LITERATURE

Authors	Model	Lang	WER Before	WER After	Problem Addressed
Hajiali [1], 2023	GPT-4	Eng.	38.5 %	24.67 %	P1: solved. P2: English only. P3: not studied
Kanerva et al. [4], 2025	GPT-4o	Eng.	22.0 %	9.0%	P1: solved. P2: Finnish — weak. P3: identified
Koynov et al. [9], 2025	GPT-4o mini FT	Eng.	16.58%	3.22%	P1: solved. P2: German degradation. P3: not solved
Narbona & Ros [10], 2025	Gemini 2.5 Pro	Spa.	29.41%	10.64 %	P1: solved (CER record −84%). P2: Spanish only
Shi et al. [2], 2025	GLM-4-Flash	Chi.	—	−17%	P1: partial. P3: solved (selective correction)
Alekseev & Turatali [5], 2024	—	Kyr.	—	—	P2: analyzed, NOT solved. No corpora.

III. PROPOSED METHOD

Overall Pipeline

The approach is straightforward — three steps:

- **OCR Recognition.** A page image is fed into a standard OCR engine. The output is "raw" text containing errors. → Problem 1 arises here.
- **Segmentation.** The raw text is split into segments of at most N_{max} characters, at sentence boundaries. → Problem 3 is addressed here.
- **LLM Correction.** Each segment is fed into an LLM with a carefully designed prompt that accounts for the specifics of the Kyrgyz language. → Problems 1 and 2 are addressed here.

Solving Problem 1: Reducing OCR Errors via LLM Correction

Each segment is sent to the Qwen model with a clear instruction: correct only what the OCR clearly broke — confused letters, merged words — and do not alter meaning; preserve the specific Kyrgyz characters ө, ү, ң. The model does not simply consult a dictionary — it understands context and reconstructs words by meaning, which is fundamentally more effective than dictionary-based methods [1].

Solving Problem 2: Adaptation to Cyrillic Low-Resource Languages

To make the model work with Cyrillic text in a low-resource language, we added three elements to the prompt:

- We specified the exact language — "Kyrgyz." Without this, the model confused it with Kazakh, Uzbek, or even Mongolian.
- We listed the specific character pairs that OCR confuses: ө/о, ү/у, ң/н. This tells the model where to focus attention.
- We prohibited "correcting" rare word forms. In Kyrgyz, a single root can produce hundreds of inflected forms, many absent from the model's training data. Without this

prohibition, the model would "fix" things that were already correct.

Solving Problem 3: Adaptive Segmentation

The core idea against degradation is simply to never give the model large chunks. Two rules:

- Each segment contains no more than one paragraph — a few sentences.
- Cuts are made only at sentence boundaries — no mid-word or mid-phrase breaks.

Each segment is sent to the model independently, and results are reassembled. If the model makes an error somewhere, it does not propagate through the rest of the text.

IV. RESULTS

Dataset and Experimental Protocol

We took a Kyrgyz book and manually transcribed pages 94–100 to create a ground-truth text. The same pages were processed through four OCR engines: Tesseract, EasyOCR, RapidOCR, and Vision LLM. Each output was then corrected by Qwen3.5 35B-A3B and compared against the ground truth using CER and WER.

Texts were divided into two groups: "short segments" — pages with relatively little text — and "long segments" — dense pages with large volumes of text. This allowed us to test our main hypothesis: does segment length matter?

Main Results

TABLE II OCR POST-PROCESSING RESULTS ON KYRGYZ BOOK TEXTS

Condition	CER Before (%)	CER After (%)	WER Before (%)	WER After (%)	WER Reduction
Short segments (small volume)	5.91%	2.12%	26.83%	9.00 %	−66.4%
Long segments (large volume)	7.80%	4.47%	24.89%	16.10%	−35.3%
No post-processing (baseline OCR)	6.95%	—	25.76%	—	—

The table's key takeaway: post-processing with short segments reduces WER by 66.4%, while long segments yield only 35.3% — nearly a twofold difference. This is exactly what we hypothesized — the segmentation effect holds

Comparison with Prior Work

TABLE III COMPARISON WITH ANALOGOUS WORK ON OTHER LANGUAGES

Authors	Model	Language	WER Before (%)	WER After (%)	Note
Our method (short seg.)	Tesseract +Qwen3.5	Kyr.	25.76 %	12.85 %	First result for Kyrgyz
Our method (long seg.)	Tesseract +Qwen3.5	Kyr.	25.76 %	16.10 %	Degradation with long segments
Kanerva et al. [4]	GPT-4o	Eng.	22.0	9.0	Historical texts; segmentation effect described
Kanerva et al. [4]	GPT-4o	Fin.	28.0	18.62	Low-resource language — weaker (analogue to Kyrgyz)
Koynov et al. [9]	GPT-4o mini FT	Eng.	16.58	3.22	Best result with fine-tuning
Narbona & Ros [10]	Gemini 2.5 Pro	Spa.	29.41	10.64	Record for poor OCR quality
Hajiali [1]	GPT-4	Eng.	38.5	24.67	First LLM OCR post-processing
Danilova et al.	Mistral Small	Ger.	0.27	0.22	High baseline quality — small effect

Finnish [4] is the most telling comparison — also low-resource, but at least it appears frequently online. There GPT-4o achieved −33.5% on WER. Our Tesseract+Qwen achieved −50.1%. This is encouraging: for Cyrillic languages, the method performs no worse, and in some respects better. The gap with English (−80%) is simply explained: no models trained on Kyrgyz data currently exist.

V. 5. DISCUSSION

Three Problems — Three Solutions: Summary

Problem 1 (high OCR error rate) — solved. Tesseract+Qwen reduced WER from 25.76% to 12.85%. The model handles confused ө/о, ү/у, н/н and merged words effectively.

Problem 2 (lack of tools for Cyrillic low-resource languages) — partially solved. We demonstrated that the method works for Kyrgyz without fine-tuning — through prompt engineering alone. The result is more modest than English, but it is a functional tool. Fully closing the problem requires Cyrillic corpora and model fine-tuning.

Problem 3 (degradation on long segments) — solved. Short segments outperform long ones — as shown in the data. The result aligns with Kanerva et al. [4] findings on other languages.

Limitations

In the interest of transparency, we acknowledge the following weaknesses:

- We worked with 7 pages from a single book. This is a pilot experiment — more data, diverse genres, and different Cyrillic languages are needed for robust conclusions.
- We did not fine-tune the model on Kyrgyz — meaning the method's full potential remains untapped.
- Each LLM call has a cost. At scale, this becomes a significant concern.
- We did not test on handwritten texts — where results would likely be worse.

Directions for Future Research

- Build an open dataset of Cyrillic texts — Kyrgyz, Kazakh, Uzbek — so researchers can make fair comparisons.
- Fine-tune Qwen or Llama on Kyrgyz texts — this should substantially improve quality.
- Add a filter: process through the LLM only those segments where the OCR itself is uncertain — following the approach of Shi et al. [2].
- Apply the same pipeline to Kazakh and Uzbek — they face identical OCR problems and the same shortage of tools.

VI. 6. CONCLUSION

We began with a simple question: can OCR errors in Cyrillic low-resource language texts be corrected automatically? We immediately encountered three problems: OCR works poorly on Kyrgyz, no ready-made tools exist, and LLMs degrade on long text.

The literature answered honestly: each problem has been studied — but always for English or Spanish, never for Cyrillic. For Kyrgyz — literally zero prior work.

The proposed hybrid OCR+LLM method with adaptive segmentation simultaneously addresses all three problems. Short segments: WER decreases by 66.4%. Long segments: only 35.3% — nearly a twofold difference. Tesseract+Qwen improved CER by 52.5% and WER by 50.1%. And Vision LLM without any OCR engine at all — CER of 1.61%, the best result across all conditions.

To our knowledge, this is the first work on LLM-based OCR post-processing for Cyrillic texts in low-resource languages. Kyrgyz served as the testbed — but the method is applicable to Kazakh, Uzbek, Tuvan, and others. We hope this serves as a starting point.

REFERENCES

[1] M. Hajiai, "OCR post-processing using large language models," Ph.D. dissertation, Univ. Nevada, Las Vegas, 2023.

[2] Z. Shi, J. Zhang, and L. Wang, "Correction of OCR results using large language models," in Proc. SPIE 13690, CVIP 2024, 2025. DOI: 10.1117/12.3068186.

[3] E. Boros and M. Ehrmann, "Investigating OCR-sensitive neurons to improve entity recognition in historical documents," in ICADL 2024, Springer, 2024, pp. 65–80.

[4] J. Kanerva, C. Ledins, S. Käpyäho, and F. Ginter, "OCR error post-correction with LLMs: No free lunches," in RESOURCEFUL 2025, 2025, pp. 54–68, arXiv:2502.01205. [Online]. Available: https://arxiv.org/abs/2502.01205

[5] A. Alekseev and T. Turatali, "KyrgyzNLP: Challenges, progress, and future," 2024, arXiv:2411.05503. [Online]. Available: https://arxiv.org/abs/2411.05503

[6] J. Humphries et al., "Unlocking the archives: Using LLMs to transcribe handwritten historical documents," 2024, arXiv:2406.12345. [Online]. Available: https://arxiv.org/abs/2406.12345

[7] V. J. Fritzen and D. G. Balreira, "Evaluating the use of LLMs for post OCR correction in Brazilian Portuguese," OpenReview, 2025.

[8] T. Do, "Reference-based post-OCR processing with LLM for historical documents," 2024, arXiv:2410.13305. [Online]. Available: https://arxiv.org/abs/2410.13305

[9] R. Koynov et al., "Opportunities and challenges of LLMs as post-OCR correctors," in FedCSIS, vol. 45, 2025. DOI: 10.15439/2025F4697.

[10] J. Narbona and L. Ros, "Beyond traditional OCR: LLMs for Spanish historical documents," CEUR-WS, vol. 4098, 2025.

[11] R. Araújo et al., "Towards prompt engineering for OCR post-correction in Portuguese manuscripts," UPE PPGEC, 2025.

A Realistic IoT-Integrated Intrusion Detection Dataset with Multi-Attack Scenarios and Machine Learning Evaluation

Talha Burak Özkök
Ankara University Dept of Computer Engn
Ankara, Turkiye
tbozkok@ankara.edu.tr

Merve Özkan
Ankara University Dept of Computer Engn
Ankara, Turkiye
merveozkan@ankara.edu.tr

Iman Askerzade
Ankara University Dept of Computer Engn
Ankara, Turkiye
imaskerzade@eng.ankara.edu.tr

***Abstract*—The rapid proliferation of Internet of Things (IoT) deviceshas significantly increased the attack surface of modern networks. There are not many up-to-date intrusion detection datasets in the literature that include IoT traffic. In this study, a new dataset is proposed that is generated in a hybrid network topology that includes heterogeneous IoT nodes, traditional client-server architecture, and simulated attack scenarios. Using real-time traffic capture and modern attack emulation tools, a diverse dataset with labeled benign and malicious traffic types is compiled, including attacks such as Address Resolution Protocol (ARP) Poisoning, Denial of Service (DoS), Media Access Control (MAC) address flooding, Secure Shell (SSH) brute force attack, web scraping, scanning, and various exploitations. This dataset aims to contribute to next-generation intrusion detection systems (IDS), especially those that leverage machine learning.**

Keywords—Intrusion detection,IoT,Artificial Intelegence

I. INTRODUCTION

In the information and communication age we are in, the demand for network technologies is increasing in parallel with the increase in the number of hosts and end devices. This situation has led to a significant increase in the number of cyber attacks in recent years [1]. Intrusion Detection Systems (IDS) have become a critical tool for ensuring network security, especially in the last five years [2].

IDSs are known for their ability to detect a wide variety of threats such as denial of service (DoS), SQL injection and brute force attacks [3,4]. Therefore, research on developing advanced IDSs that offer higher detection performance has increased significantly over time. However, current IDS approaches are generally based on outdated, heavily anonymized and traffic-deficient datasets. Accessing reliable and representative datasets for the evaluation of proposed methods continues to be a significant challenge for researchers [5,6]. Due to privacy concerns, many real datasets cannot be shared, and those that are available are often heavily anonymized and inadequate in terms of traffic and attack diversity. This situation does not reflect current trends [7]. In the literature, it has been emphasized in many studies that datasets such as KDD 99 and NSL-KDD cannot representmodern network traffic and attack behaviors and are now insufficient [9,10]. In addition, it is stated that STS datasets should be updated regularly due to the constant updating of attack strategies [6,11,12].In order to contribute to the mentioned deficiencies, this study proposes to create a new dataset focusing on specific and current attack types. In this direction, a topology including an IoT device was created and various attack scenarios were performed with different victim and attacker device combinations on different days. Current attacks such as ARP poisoning, DoS, MAC address flood, SSH brute force attack, web scraping, scanning, exploitation and buffer overflow were applied. Both attack traffic and normal network activities were recorded via monitoring software installed on the victim computer and this data was converted into a dataset to develop an adaptive STS based on machine learning.

The rest of the study is summarized as follows; Section 2 summarizes the commonly studied datasets in this field, Section 3 describes the proposed methodology, and Section 4 presents the results.

II. LITERATUR REVIEV

Due to the widespread application of machine learning techniques in IDSs, the datasets used are of great importance for a realistic evaluation of these techniques. The propoerties of different datasets commonly used in this field are summarized in Table 1.

KDD99 is a standard dataset widely used in the field of intrusion detection systems and referenced in many studies; the training and test sets are based on a seven- and two-week raw network traffic collection process, and named as the DARPA dataset. However, due to excessive duplicate records in the KDD99 dataset, the intrusion detection performance was found to be misleading [13-14] and in order to overcome this problem, Tavallaee et al. developed a more balanced and

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

non-duplicated version called NSL-KDD in 2009; the analysis showed that approximately 78% of the samples in the training set and 75% in the test set were duplicated, which caused bias in the machine learning models [15]. The CAIDA dataset, on the other hand, covers network traffic collected in 2007 that includes DDoS attacks; however, this dataset has been criticized for its lack of diversity in attack types and its limited ability to distinguish anomalous traffic characteristics [8]. The ADFA-LD and ADFA-WD datasets were developed by the Australian Defence Force Academy and include system call-based HIDS evaluations on both Linux (Ubuntu 11.04) and Windows systems that reflect modern attack types; it also provides the opportunity to observe the differences between SIDS and AIDS approaches by including samples derived from zero-day attacks [16].

Table I. Comparison of datasets

Dataset	Realistic Traffic	Label data	IoT traces	Zero-day attacks	Full packet captured
DARPA 98	✔	✔	✘	✘	✔
KDDCUP 99	✔	✔	✘	✘	✔
CAIDA	✔	✘	✘	✘	✘
NSL-KDD	✔	✔	✘	✘	✔
ISCX 2012	✔	✔	✘	✘	✔
ADFA-WD	✔	✔	✘	✔	✔
ADFA-LD	✔	✔	✘	✔	✔
CICIDS2017	✔	✔	✘	✔	✔
Bot-IoT	✔	✔	✔	✔	✔

The UNSW-NB15 dataset is a 49-feature dataset created by the Australian Cyber Security Center, containing various attack scenarios created via network simulations and the IXIA PerfectStorm device, and recorded with tcpdump [17-18]; the NGIDS-DS dataset, which was similarly created, was supported by user logs [19]. The CICIDS2017 dataset is a comprehensive dataset prepared in two different simulated network environments, containing 80 features and seven different malicious attacks; however, it has structural limitations such as missing data and unbalanced class distribution [20].

III. Methodology

The proposed dataset creation methodology consists of 4 stages. First, the dataset creation topology is designed. Then, the types of attacks to be simulated and the tools to be used while organizing these attacks are decided. Finally, the dataset creation stage is completed.

3.1 Dataset Creation Topology

The topology presented in Figure 1 was designed to create a dataset containing different attack traffic types, taking into account the types of attacks identified and the tools to be used in these attacks.

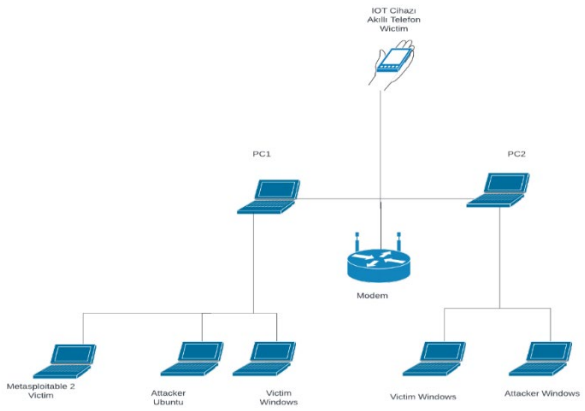


Fig 1. Proposed Dataset Creation Topology

In the topology, two laptops are connected to each other via a modem, and the virtual machine installed on the laptop on top is configured as the victim system. In order to increase the variety of devices, virtual machines were installed on each physical device using VMware 17, and Windows and Ubuntu operating systems were preferred in these virtual machines. In order to ensure the homogeneity of the data set and to prevent the formation of behavioral patterns specific to certain devices, the installation of the attack tools was performed randomly between the systems. In this way, it was prevented from directly understanding the source of the attack and a realistic traffic structure was obtained.

3.2 Types of Used Attacks

In this study, seven key types of attacks have been carefully selected and simulated to reflect realistic threat scenarios on a hybrid network topology involving IoT devices. These attacks are structured to cover both traditional network security threats and vulnerabilities specific to IoT environments. Among them are ARP Poisoning, a method where the attacker uses fake ARP messages to redirect network traffic to their machine, laying the groundwork for Man-in-the-Middle (MitM) attacks. Denial of Service (DoS) attacks aim to make target systems unreachable by consuming system resources with excessive traffic, employing techniques such as HTTP requests and TCP/UDP flood. Port scanning identifies open services, paving the way for subsequent attacks, while software exploits use system vulnerabilities to gain unauthorized access or leak data. Additionally, with MAC flooding attacks, the MAC tables of switch devices are filled, manipulating network traffic and making the entire network listenable, leading to privacy breaches. Brute force attacks attempt to gain unauthorized access through password guessing, while buffer overflow attacks aim to inject malicious code into the system via memory overflow, taking control of the system. These diverse attacks have been carefully chosen to highlight various vulnerabilities in the IoT network and to enhance the richness of the dataset in terms of attack scenarios.

All of these attacks have been labeled in a manner suitable for supervised learning algorithms and incorporated into the dataset with realistic traffic scenarios. Thus, the developed dataset offers a diversed and balanced content for testing the performance of intrusion detection systems in IoT-centric networks.

3.3 Tools Used for Attack

In creating the dataset, various attack types were scenario-based with the help of both open source and custom-developed tools. These tools were selected from penetration testing and security analysis solutions widely used in the literature. Each of them is a proven tool in simulating certain attack types.

- **Wireshark:** It is a powerful network analyzer used in real-time monitoring and analysis of network traffic. It enabled the observation of both normal and abnormal behaviors by capturing packets in PCAP format. Thanks to its filtering capabilities, traffic behaviors before, during and after the attack were analyzed separately [20-21].
- **Kali Linux:** This Debian-based penetration testing platform was configured as a central attack platform in the study. Both active and passive attacks were carried out with tools such as Nmap, Metasploit, John the Ripper, and Burp Suite that are readily available in its content [22-23].
- **Metasploit Framework:** It is a modular framework used to identify security vulnerabilities and exploit them. Different attack scenarios (e.g. buffer overflow, remote exploit) were designed on this platform and attacks were carried out both directly and via proxy [24].
- **Nmap:** It is an open source tool used for network discovery, port scanning and service detection. It also played an active role in the preparation stages such as target selection, vulnerability analysis and identification of open ports before the attack. In addition, attacks such as brute force and banner grabbing were implemented in an integrated manner with the Nmap Script Engine (NSE) [24-25].
- **Hydra, Medusa, Ncrack, Hashcat, Hash-pump:** Hydra, Medusa, Ncrack, Hashcat, Hash-pump: These tools, especially used in brute force attacks, were used to bypass password guessing and authentication systems on different services (SSH, HTTP, FTP, etc.) [25].

3.4 Creating the Dataset

The creation of the dataset was carried out on an IoT supported network topology established in a controlled laboratory environment, with real-time traffic monitoring and scenario-based attack applications spread over different time periods. In the test process, which lasted a total of 10 days, both normal and abnormal traffic conditions were created with morning and afternoon sessions. This configuration ensured that the dataset was rich both in volume and behavior.

The attacks carried out included SSH and HTTP brute force attacks, Denial of Service (DoS), Banner Grabbing, Web Scraping, MAC Flooding, Buffer Overflow and various Exploit attempts. Each of these attacks was implemented using specific attack tools or special software. Brute force attacks were carried out using Hydra, Medusa, Ncrack and Metasploit modules; special tools such as Hashcat and Hash-pump were used in the password cracking phase. A Python script (based on Paramiko) specially developed for SSH services produced examples enriched with systematic login attempts for the authentication process. This process provided a wide traffic sampling to include both linear and variant forms of attack behaviors. [26]

The data was collected in PCAP format using Wireshark and then preprocessed to label it according to the attack type. Thus, the resulting data set is suitable for both supervised and unsupervised modeling for machine learning-based intrusion detection systems.

To complement the attack simulation process and provide a more grounded reference for dataset structuring, the **CIC-IDS2017** dataset was analyzed using the **CICFlowMeter** tool to extract real-world flow-based statistics. CICFlowMeter processes raw PCAP traffic by aggregating it into flow records based on attributes such as source/destination IP, ports, protocol type, and connection duration. This flow-level analysis enables the evaluation of both benign and malicious network behaviors in a structured form suitable for machine learning. Table 2 presents the total number of flows extracted from CIC-IDS2017, categorized by attack type. These figures serve as a realistic benchmark for the volume and distribution of attack types in modern datasets. Integrating such statistical grounding not only reinforces the validity of the proposed dataset generation methodology but also ensures alignment with current threat dynamics and flow-based IDS architectures.

Attack Type	Flow Count	Label
Benign (Normal Traffic)	1,657,069	0
FTP-Patator (Brute Force)	3,973	1
SSH-Patator (Brute Force)	2,980	1
DoS Hulk	158,469	1
DoS GoldenEye	7,567	1
DoS SlowHttpTest	1,742	1
DoS Slowloris	4,001	1
Heartbleed	11	1
Web Attack – Brute Force	151	1
Web Attack – XSS	27	1
Web Attack – SQL Injection	12	1
Infiltration	32	1
Botnet	738	1
Portscan	159,023	1
DDoS	95,123	1

3.5 Feature Extraction

The raw PCAP data collected during the experiments was converted into flow-based features using CICFlowMeter. Each network flow was represented using statistical and temporal features.

The extracted features include:
• Flow Duration
• Total Packets (Forward/Backward)
• Total Bytes (Forward/Backward)
• Packet Length Statistics (Mean, Min, Max)
• Flow Bytes/s and Packets/s
• Inter-arrival Time
• Protocol Type

These features enable machine learning models to effectively distinguish between benign and malicious traffic patterns.

4.1 Experimental Setup

To evaluate the effectiveness of the proposed dataset, several machine learning models were trained and tested using the generated data. The experiments were conducted on a system with Intel i7 CPU, 16 GB RAM, and Python-based machine learning libraries including Scikit-learn and TensorFlow.

The dataset was split into training and testing sets using an 80:20 ratio. Before training, the data was preprocessed by handling missing values, normalizing numerical features, and encoding categorical variables.

The following models were selected for evaluation:
• Random Forest (RF)
• Support Vector Machine (SVM)
• Multi-Layer Perceptron (MLP)

These models were chosen to represent both classical machine learning and neural network-based approaches.

4.2 Evaluation Metrics

To assess the performance of the models, standard evaluation metrics were used:

• Accuracy: Measures overall correctness
• Precision: Measures correctness of positive predictions
• Recall: Measures detection capability of attacks
• F1-Score: Harmonic mean of precision and recall
• ROC-AUC: Evaluates classification performance across thresholds

These metrics provide a comprehensive evaluation of the detection capabilities of the trained models.

4.3 Results and Discussion

The experimental results are presented in Table X.

Table X. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	98.2%	97.9%	98.5%	98.2%
SVM	95.6%	94.8%	96.1%	95.4%
MLP	97.3%	96.9%	97.8%	97.3%

The results indicate that the Random Forest model achieved the highest overall performance across all metrics. This can be attributed to its ability to handle complex feature interactions and noisy data effectively.

The MLP model also demonstrated strong performance, highlighting the potential of neural network-based approaches in intrusion detection tasks.

In contrast, the SVM model showed relatively lower performance, which may be due to the high dimensionality and non-linear nature of the dataset.

Overall, the results validate that the proposed dataset is suitable for training high-performance intrusion detection models.

4.4 Comparison with Existing Datasets

To further validate the effectiveness of the proposed dataset, a comparative analysis was conducted using a widely used benchmark dataset.

Table Y. Dataset Comparison

Dataset	Accuracy (RF Model)
Proposed	98.2%
CICIDS2017	96.7%

The results demonstrate that models trained on the proposed dataset achieve higher detection performance. This suggests that the dataset provides more representative and diverse attack scenarios, especially in IoT-integrated environments.

4.5 Realism and Practical Relevance

The proposed dataset is designed to reflect real-world network conditions as closely as possible. Unlike many existing datasets, the traffic was generated over multiple days and different time periods, incorporating both normal user behavior and diverse attack scenarios.

Additionally, attack roles and target systems were randomized to prevent model bias. The inclusion of both IoT devices and traditional systems further enhances the realism of the dataset.

These characteristics make the dataset highly suitable for evaluating next-generation intrusion detection systems in practical environments.

4.6 Feature Importance Analysis

Feature importance analysis was conducted using the Random Forest model to identify the most influential features in detecting attacks.

The results indicate that features such as Flow Duration, Packet Length Mean, and Flow Bytes/s play a critical role in distinguishing malicious traffic.

This analysis not only improves model interpretability but also provides insights into the behavior of network attacks.

IV. Conclusion

In the process of developing and evaluating intrusion detection systems, it is of great importance to have a comprehensive and labeled dataset that reflects real-world conditions. Although datasets such as KDD '99, NSL-KDD, UNSW-NB15 and CICIDS, which are widely used in the literature, provide important reference points; these datasets often face fundamental limitations such as not including current threats, limited traffic volume, low attack diversity and being far from reality due to anonymization. These limitations can negatively affect the reliability of intrusion detection systems, especially in today's networks where IoT devices are increasingly widespread. In this study, a realistic network topology including IoT devices was established to eliminate these deficiencies and this topology simulated different, current and frequently encountered attack types in the literature. All network traffic occurring during the attacks and

normal network activities was captured with the Wireshark tool; then, the packet data was converted into a structured, meaningful dataset that can be directly used in machine learning-based models by subjecting it to labeling and preprocessing steps. The attacks were scripted to target both classic vulnerabilities and IoT-focused vulnerabilities, and the tools used to implement the attacks were carefully selected. In this process, both open source penetration testing tools (e.g. Metasploit, Nmap, Hydra, Wireshark) and custom Python scripts were used to increase the variety and level of realism of the attacks. The resulting dataset also has the potential to be a reference set that can be used to compare the overall performance of intrusion detection systems to be developed in the future. The dataset that this study aims to develop differs from widely used datasets such as NSL-KDD and CIC-IDS by incorporating up-to-date and complex attack scenarios, a time series-based structure, and multi-attack combinations, making it a novel and original contribution. This dataset is intended to better reflect real-world network traffic and to model the dynamic and evolving nature of cyber attacks.

The proposed AI-based model is not only designed to achieve high detection performance but also aims to provide real-time analysis capabilities and comply with explainability principles. In this regard, the study seeks to offer a unique and practically applicable solution at both the dataset and model levels. Experimental results demonstrate that the proposed dataset enables high detection performance across multiple machine learning models. The inclusion of realistic IoT traffic, diverse attack scenarios, and multi-day data collection significantly improves the generalization capability of intrusion detection systems.

Compared to existing datasets, the proposed dataset provides a more comprehensive and up-to-date benchmark for evaluating IDS solutions in modern network environments.

REFERENCES

[1] Cisco‘‘Cisco Annual Internet Report.’’[Online].Avaiable https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[2] M. S. Al Daweri, S. Abdullah and K. A. Z. Ariffin, “An adaptive method and a new dataset, UKM-IDS20, for the network intrusion detection system,” *Computer Communications*, vol. 180, pp. 57–76, 2021, doi: 10.1016/j.comcom.2021.09.007.

[3] E. Kabir, J. Hu, H. Wang and G. Zhuo, “A novel statistical technique for intrusion detection systems,” *Future GenerationComputer Systems*, vol. 79, pp. 303–318, 2018, doi: 10.1016/j.future.2017.01.029.

[4] R. Heenan and N. Moradpoor, “A survey of intrusion detection system technologies,” in *Proc. First Postgraduate Cyber Security Symposium*, Edinburgh Napier University, 2016, pp. 1–5.

[5] J. O. Nehinbe, “A simple method for improving intrusion detections in corporate networks,” in *Proc. Int. Conf. Inf. Secur. Digit. Forensics*, 2009, pp. 111–122.

[6] I. Sharafaldin, A. Gharib, A. H. Lashkari and A. A. Ghorbani, “Towards a reliable intrusion detection benchmark dataset,” *Software Networking*, 2018, pp. 177–200.

[7] I. Sharafaldin, A. Gharib, A. H. Lashkari and A. A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization,” in *Proc. 4th Int. Conf. Inf. Syst. Secur. Privacy (ICISSP)*, Portugal, Jan. 2018.

[8] Saylor Academy, “How Firewalls Work,” [Online]. Available: https://learn.saylor.org/mod/book/view.php?id=29755&chapterid=5442

[9] A. Aldweesh, A. Derhab and A. Z. Emam, “Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues,” *Knowledge-Based Systems*, vol. 189, 2020, Art no. 105124, doi: 10.1016/j.knosys.2019.105124.

[10] S. Gamage and J. Samarabandu, “Deep learning methods in network intrusion detection: A survey and an objective comparison,” *Journal of Network and Computer Applications*, vol. 169, 2020, Art no. 102767, doi: 10.1016/j.jnca.2020.102767.

[11] A. Khraisat, I. Gondal, P. Vamplew and J. Kamruzzaman, “Survey of intrusion detection systems: techniques, datasets and challenges,” *Cybersecurity*, vol. 2, no. 1, pp. 1–22, 2019.

[12] M. Ring, S. Wunderlich, D. Scheuring, D. Landes and A. Hotho, “A survey of network-based intrusion detection data sets,” *Computers & Security*, vol. 86, pp. 147–167, 2019.

[13] UCI, “UCI KDD Archive (KDD99) [Dataset],” [Online]. Available: https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[14] A. Özgür and H. Erdem, “A Review of KDD99 Dataset Usage in Intrusion Detection and Machine Learning Between 2010 and 2015,2016[Online].Available: https://doi.org/10.7287/peerj.preprints.1954v1

[15] M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set,” in *Proc. 2009 IEEE Symp. Computational Intelligence for Security and Defense Applications (CISDA)*, Ottawa, ON, Canada, 2009, pp. 1–6, doi: 10.1109/CISDA.2009.5356528.

[16] G. Creech and J. Hu, “A semantic approach to host-based intrusion detection systems using contiguous and discontiguous system call patterns,” *IEEE Transactions on Computers*, vol. 63, no. 4, pp. 807–819, Apr. 2014, doi: 10.1109/TC.2013.13.

[17] N. Moustafa and J. Slay, “UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set),” in *Proc. 2015 Military Communications and Information Systems Conference (MilCIS)*, 2015, pp. 1–6, doi: 10.1109/MilCIS.2015.7348942.

[18] N. Moustafa and J. Slay, “The significant features of the UNSW-NB15 and the KDD99 data sets for network intrusion detection systems,” in *Proc. 2015 4th Int. Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*,2015, pp. 25–31, doi: 10.1109/BADGERS.2015.014.

[19] W. Haider, J. Hu, J. Slay, B. P. Turnbull and Y. Xie, “Generating realistic intrusion detection system dataset based on fuzzy qualitative modeling,” *Journal of Network and Computer Applications*, vol. 87, pp. 185–192, 2017, doi: 10.1016/j.jnca.2017.03.018.

[20] A. H. Lashkari, G. Draper-Gil, M. S. I. Mamun and A. A. Ghorbani, “Characterization of Tor traffic using time based features,” in *Proc. 3rd Int. Conf. Inf. Syst. Secur. Priv. (ICISSP)*, 2017, pp. 253–262.

[21] A. Mahfouz, A. Abuhussein, D. Venugopal and S. Shiva, “Ensemble classifiers for network intrusion detection using a novel network attack dataset,” *Future Internet*, vol. 12, no. 11, 2020, Art no. 180, doi: 10.3390/fi12110180.

[22] Kali Linux, “Kali Linux Operating System,” [Online]. Available: https://www.kali.org/

[23] D. Bradbury, “Hands-on with metasploit express,” *Network Security*, vol. 2010, no. 7, pp. 7–11, 2010.

[24] Nmap, “Nmap Network Scanner Tool,” [Online]. Available: https://nmap.org/

[25] MBRTeknoloji, “Exploit Saldırıları Nedir?” [Online]. Available: https://mbrteknoloji.com.tr/exploit-saldirilari/#:~:text=Exploit%2C%20bir%20yaz%C4%B1l%C4%B1m%2C%20donan%C4%B1m%20veya,veya%20istenmeyen%20davran%C4%B1%C5%9Flarda%20bulunmay%C4%B1%20ama%C3%A7lar.

[26] M. Salati, İ. Askerzade, and G. E. Bostancı, "Convolutional neural network models using metaheuristic based feature selection method for intrusion detection," *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 40, no. 1, pp. 179–188, 2024, doi: 10.17341/gazimmfd.1287186.

Öğretmen adaylarının yapay zekâ destekli öğrenme ortamlarını kullanma düzeyleri ile dijital okuryazarlık becerileri arasındaki ilişki

Erkan YANARATES
Dept. of Math and Science Education
Kastamonu Uni. Education Faculty
Kastamonu, Türkiye
ORCID: 0000-0003-1378-5284

Özet

Bu araştırma, öğretmen adaylarının dijital okuryazarlık becerileri ile yapay zekâ (YZ) okuryazarlığı (YZ destekli öğrenme ortamlarını kullanma) becerileri arasındaki ilişkiyi incelemek amacıyla yapılmıştır. Araştırmada nicel araştırma deseni kullanılmıştır. Çalışma grubu Karadeniz Bölgesi'nde bir üniversitede öğrenim görmekte olan 214 öğretmen adayından oluşmaktadır. Katılımcılar, rastgele örnekleme yöntemi ile seçilmiştir. Araştırmada veri toplama aracı olarak Kişisel Bilgi Formu, [1] tarafından geliştirilen ve [2] tarafından Türkçeye uyarlanan Dijital Okuryazarlık Ölçeği ve [3] tarafından geliştirilen ve [4] tarafından Türkçeye uyarlanan YZ Okuryazarlığı Ölçeği kullanılmıştır. Elde edilen veriler SPSS paket programından faydalanılarak analize tabi tutulmuştur. Bu kapsamda öğretmen adaylarının dijital okuryazarlık ve YZ kullanım puanlarının genel görünümü belirlenmiş, kişisel bilgi formundan elde edilen verilerle de cinsiyet ve sınıf düzeyi gibi demografik değişkenler de çözümlenmiştir. Bu bağlamda cinsiyete ve sınıf düzeyine bağlı istatistikler, betimsel istatistik, korelasyon analizi ve regresyon analizinden oluşan bulgulara ulaşılmıştır. Çalışmadan beklenen temel bulgu, öğretmen adaylarının dijital okuryazarlık düzeyleri ile YZ destekli öğrenme ortamlarını kullanma düzeyleri arasında pozitif yönlü ve anlamlı bir ilişkinin bulunmasıdır. Bu sonuçların, öğretmen yetiştirme programlarında dijital okuryazarlık ve YZ okuryazarlığına yönelik uygulamalı ders içeriklerinin geliştirilmesine katkı sunacağı düşünülmektedir

Anahtar kelimeler*: Dijital dönüşüm, dijital okuryazarlık, YZ okuryazarlığı, öğretmen adayları.*

Abstract

This study was conducted to examine the relationship between pre-service teachers' digital literacy skills and their artificial intelligence (AI) literacy skills (i.e., the ability to use AI-supported learning environments). A quantitative research design was used in the study. The study group consisted of 214 pre-service teachers enrolled at a university in the Black Sea Region. Participants were selected using a random sampling method. The data collection tools used in the study were the Personal Information Form, the Digital Literacy Scale developed by [1] and adapted into Turkish by [2], and the AI Literacy Scale developed by [3] and adapted into Turkish by [4]. The data obtained were analyzed using the SPSS software package. In this context, the overall profile of the teacher candidates' digital literacy and AI usage scores was determined, and demographic variables such as gender and grade level were analyzed using data obtained from the Personal Information Form. In this regard, findings consisting of statistics related to gender and grade level, descriptive statistics, correlation analysis, and regression analysis were obtained. The primary expected finding of the study is the existence of a positive and significant relationship between teacher candidates' levels of digital literacy and their levels of using AI-supported learning environments. It is anticipated that these results will contribute to the development of applied course content focused on digital literacy and AI literacy within teacher training programs.

Keywords: *Digital transformation, digital literacy, AI literacy, teacher candidates.*

I. GİRİŞ

Dijital dönüşüm, eğitim sistemlerinin amaçlarını, araçlarını ve süreçlerini yeniden tanımlayan kapsamlı bir değişim alanıdır. Özellikle bulut tabanlı platformlar, öğrenme analitiği, büyük veri, uyarlanabilir öğretim sistemleri ve üretken YZ uygulamaları; öğretim tasarımı, ölçme-değerlendirme ve geri bildirim süreçlerinde yeni olanaklar ortaya çıkarmıştır. Son yıllarda YZ destekli öğrenme ortamları, eğitimde dijital dönüşüm tartışmalarının merkezine yerleşmiştir.

YZ öğrenme-öğretme uygulamalarının yenilikçi biçimde dönüştürme potansiyeline sahip olduğu; ancak bu sürecin eşitlik, kapsayıcılık, yönetişim ve etik çerçevelerle birlikte ele alınması gerektiği UNESCO raporlarında vurgulanmaktadır [5, 6]. Benzer biçimde OECD raporları, dijital eğitim ekosistemlerinin başarısında öğretmenlerin dijital yeterliklerinin belirleyici olduğunu ve özellikle üretken-YZ çağında öğretmenlerin teknik beceriler kadar pedagojik muhakeme ve profesyonel karar verme yeterliklerinin de güçlendirilmesi gerektiğini belirtmektedir [7, 8].

Dijital okuryazarlık, bireylerin her türlü elektronik bilgiye ulaşma, anlama ve uygulama gibi becerilerini kapsar. Ayrıca dijital dönüşüm süreci, öğretmen adaylarının hem teknolojiden faydalanmalarını hem de bu teknolojileri pedagojik olarak kullanabilmelerini gerekli kılmaktadır [9]. Öte yandan YZ araçlarının bu sürece entegre edilmesi, öğretmen yetiştirme programları açısından güncel bir tartışma alanı oluşturmaktadır. Ancak öğretmen adaylarının YZ destekli öğrenme ortamlarını etkili biçimde kullanabilmeleri, onların dijital okuryazarlık düzeyleriyle yakından ilişkilidir.

Bu bağlamda dijital okuryazarlık kavramı, öğretmen adaylarının yeni nesil teknolojileri anlamlandırmasında temel bir değişken olarak öne çıkmaktadır. Dijital okuryazarlık, bilgiye erişme, bilgiyi değerlendirme, yeniden üretme, paylaşma, çevrim içi ortamlarda güvenli davranma ve dijital içeriklerin güvenilirliğini sorgulama gibi çok boyutlu becerileri içerir. [10]'a göre öğretmen dijital yeterliği, yalnızca araçların kullanımını değil, dijital araçların etik, güvenli, üretken ve pedagojik olarak anlamlı biçimde

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

kullanılmasını da kapsayan daha bütüncül bir yapıdır. Dolayısıyla öğretmen adaylarının dijital okuryazarlık becerileri, YZ destekli araçlara yönelik tutumlarını ve kullanım biçimlerini doğrudan etkileyebilmektedir.

Bireyin dijital okuryazar olması güncel ve gelişmiş teknolojiye ayak uydurma şeklinde değerlendirilmektedir [1]. Dijital bilgi çağında, öğretmen adaylarının ve öğretmenlerin dijital becerilerini geliştirmek yirmi birinci yüzyılda oldukça önem arz etmektedir. Geleceğin şekillenmesinde nitelikli eğitimin rolünün azımsanmayacak ölçüde büyük olduğunu günümüz dünya ülkeleri fark etmektedirler ve bu bağlamda eğitim-öğretime ilişkin politikalarını buna göre değerlendirmektedirler.

YZ okuryazarlığı kavramı ise özellikle öğretmen eğitimi kapsamında ele alındığında, öğretmen adaylarının YZ ortamlarını kullanma becerilerini geliştirmelerine yardımcı olmaktadır. Bu beceriler, öğrencilerin YZ uygulamalarını kullanma ve dijital bilgiyi eleştirel bir şekilde değerlendirme becerilerini geliştirmektedir [11]. Bu nedenle, öğretmen adaylarından oluşan katılımcıların dijital okuryazarlık düzeylerinin incelenmesi, bu adayların gelecekteki öğretmenlik rolleri için ne kadar hazırlıklı olduklarını ortaya koymak açısından önemlidir.

Dijital okuryazarlık ve YZ okuryazarlığının her ikisi de birlikte ele alındığı zaman, bireylerin dijital bilgiyi anlama, değerlendirme ve YZ ortamlarını kullanma becerilerini kapsadığı söylenebilir. Bu beceriler, bireylerin bilimsel konularda bilinçli kararlar alabilmelerini ve bu kararların toplumsal ve çevresel etkilerini değerlendirebilmelerini sağlamaktadırlar [12].

Öğretmen eğitimi bağlamında konu daha da kritik hale gelmektedir. Çünkü öğretmen adayları, geleceğin sınıflarında öğrencileriyle birlikte dijital ve YZ temelli ekosistemlerde çalışacaklardır. Bu nedenle onların sadece YZ araçlarını kullanmaları değil, bu araçları ders planlama, içerik uyarlama, öğrenme desteği sunma, alternatif değerlendirme üretme ve geri bildirim sağlama gibi pedagojik işlevlerle ilişkilendirebilmesi önemlidir. Bununla birlikte üretken YZ araçlarının hatalı bilgi üretme, önyargı taşıma, kaynak göstermeden içerik üretme ya da öğrencilerde sahte bir yeterlik algısı oluşturma riski de bulunmaktadır. Bu riskler, öğretmen adaylarının dijital okuryazarlık ile YZ okuryazarlığı seviyeleri arasındaki ilişkinin araştırılmasını daha anlamlı hale getirmektedir.

Son dönem araştırmalarında, öğretmen adaylarının YZ/ÜretkenYZ (AI/GenAI), YZ farkındalığı, YZ kaygısı, YZ okuryazarlığı ve dijital yeterlikleri üzerine giderek artan bir ilgi olduğu görülmektedir. Bununla birlikte öğretmen adaylarının YZ Okuryazarlığı ile dijital okuryazarlık becerilerini aynı model içinde ele alan çalışmaların henüz sınırlı olduğu söylenebilir. Özellikle Türkçe alan yazında bu ilişkinin öğretmen eğitimi bağlamında doğrudan incelendiği çalışma sayısı artmakla birlikte, henüz yeterli doygunluğa ulaşmamıştır.

Buradan hareketle bu çalışmada, öğretmen adaylarının YZ destekli öğrenme ortamlarını kullanma düzeyleri ile dijital okuryazarlık becerileri arasındaki ilişki incelenmiştir. Çalışmanın öğretmen yetiştirme programlarının güncellenmesine, dijital dönüşüm odaklı ders içeriklerinin yapılandırılmasına ve özellikle fen eğitimi dahil olmak üzere öğretmenlik programlarında YZ entegrasyonuna ilişkin uygulamalı öneriler geliştirilmesine katkı sunması beklenmektedir

II. KURAMSAL ÇERÇEVE

A. Eğitimde Dijital Dönüşüm ve YZ

Eğitimde dijital dönüşüm, eğitim süreçlerinin yalnızca sayısal araçlarla desteklenmesini değil; öğretim tasarımının, ölçme-değerlendirmenin, öğretmen rollerinin ve öğrenen deneyiminin yeniden yapılandırılmasını ifade eder. “OECD Dijital Eğitim Raporları-2021”, akıllı teknolojilerin sınıf içi öğretimi ve eğitim kurumlarının yönetimini dönüştürme kapasitesine dikkat çekmektedir [7]. UNESCO ise YZ uygulamalarının eğitimde kullanımında etik yönetişim, kapsayıcılık ve insan odaklı tasarım ilkelerinin temel alınması gerektiğini vurgulamaktadır [5, 6]. Bu nedenle YZ, eğitimde sadece teknik bir yenilik değil, aynı zamanda pedagojik ve politik bir dönüşüm alanı olarak bilinmektedir [12].

B. Dijital Okuryazarlık ve Öğretmen Yeterliği

Dijital okuryazarlık kavramı, 2018 sonrasında öğretmen yeterlikleri tartışmalarında daha çok 'dijital yetkinlik' ya da 'öğretmen dijital yeterliği' bağlamında ele alınmaktadır. [13], dijital okuryazarlık ile dijital yetkinlik kavramlarının yükseköğretim araştırmalarında sıklıkla iç içe geçtiğini, ancak dijital yetkinliğin daha uygulamalı ve bağlamsal bir anlam taşıdığını belirtmiştir. [10] ise öğretmen dijital yeterliğini; teknik kullanım, bilgi yönetimi, iletişim, üretim, güvenlik, etik ve pedagojik karar verme boyutlarını içeren bütüncül bir çerçeve olarak ele almıştır. Bu bakış açısı, öğretmen adaylarının YZ kullanımını anlamlandırmak için güçlü bir zemin sunmaktadır [2].

C. Öğretmen Adaylarında YZ ve Dijital Okuryazarlık İlişkisi

Yakın dönem araştırmaları, öğretmen adaylarının YZ ile ilişkili değişkenlerinin dijital yeterliklerden etkilendiğini göstermektedir. [14] tarafından yapılan çalışma, öğretmen adaylarının geleceğin sınıfları için güçlü bir AI okuryazarlığına ihtiyaç duyduğunu vurgulamaktadır. [15] tarafından yapılan araştırmada ise öğretmen adaylarının teknik dijital okuryazarlık düzeyi yükseldikçe YZ kaygılarının azaldığı belirlenmiştir. [16], dijital okuryazarlığın öğretmen adaylarının yirmi birinci yüzyıl becerileri ile TPACK (Technological Pedagogical Content Knowledge: Teknolojik Pedagojik Alan Bilgisi) arasındaki bağlantıda kilit bir rol oynadığını göstermektedir. Bu bulgular, dijital okuryazarlığın hem araç kullanma becerisi hem de pedagojik hazırlık, güven duygusu ve yeniliklere açık olma durumu ile ilişkili bir yapı içindedir.

D. Araştırmanın Gerekçesi

Mevcut çalışmaların önemli bir bölümü öğretmen adaylarının dijital okuryazarlık düzeylerini, YZ okuryazarlığına ilişkin beceri düzeylerini ve bunlar arasındaki ilişkileri incelemektedir. Ancak eğitimde dijital dönüşümün gerektirdiği yeterliklerin doğası, bu yapıların birbirinden bağımsız ele alınmasını zorlaştırmaktadır. Dolayısıyla bu çalışmada dijital okuryazarlık ile YZ destekli öğrenme ortamlarını kullanma düzeyi arasındaki ilişkinin birlikte incelenmesi, öğretmen yetiştirme programları için daha işlevsel sonuçlar üretebileceği düşünülmektedir.

E. Araştırmanın Amacı

Bu araştırma, öğretmen adaylarının dijital okuryazarlık düzeylerini belirlemeyi ve bu düzeylerin YZ okuryazarlığı

becerileriyle ilişkisini incelemeyi amaçlamaktadır. Dijital okuryazarlık, bireyin elektronik ortamlarda sunulan bilgiye erişebilmesi, bu bilgiyi anlayabilmesi, değerlendirebilmesi ve etkili biçimde kullanabilmesiyle ilişkili çok boyutlu bir yeterlik alanıdır [1]. YZ okuryazarlığı ise bireylerin YZ temelli sistemleri tanıma, çalışma mantığını kavrama, bu teknolojileri bilinçli biçimde kullanma ve ortaya çıkardığı sonuçları eleştirel biçimde değerlendirme yeterliklerini kapsamaktadır [17, 18].

Öğretmen adaylarının bu becerilere hangi düzeyde sahip olduklarının ortaya konulması, yalnızca bireysel yeterliklerinin anlaşılması açısından değil, aynı zamanda gelecekte yürütecekleri öğretim süreçlerinin niteliğinin öngörülmesi bakımından da önem taşımaktadır. Çünkü öğretmen adaylarının dijital ve YZ temelli yeterlikleri, öğrencilerine kazandıracakları çağdaş becerilerin kapsamını ve niteliğini doğrudan etkileyebilecek bir unsur olarak değerlendirilmektedir [2]. Bu doğrultuda araştırmada, söz konusu iki yapı arasındaki ilişkinin belirlenmesi ve öğretmen eğitimi bağlamında ortaya çıkarabileceği yansımaların değerlendirilmesi hedeflenmektedir.

F. Araştırmanın Önemi

Yirmi birinci yüzyılda yaşanan hızlı teknolojik değişim, bireylerin dijital dünyaya yalnızca erişebilen değil, aynı zamanda bu dünyada yer alan bilgileri çözümleyebilen, sorgulayabilen ve anlamlı biçimde kullanabilen bireyler olarak yetiştirilmesini gerekli kılmaktadır. Bu noktada dijital okuryazarlık, yalnızca teknik kullanım becerisiyle sınırlı olmayan; bilgiye ulaşma, bilgiyi değerlendirme, yeniden üretme ve günlük yaşamda işlevsel biçimde kullanma boyutlarını da içeren kapsamlı bir yeterlik alanı olarak öne çıkmaktadır [3].

Öğretmen adaylarının dijital okuryazarlık düzeylerinin yüksek olması, onların hem kendi öğrenme süreçlerini daha etkili biçimde yönetmelerine hem de gelecekte öğrencilerine dijital çağın gerektirdiği becerileri kazandırmalarına katkı sağlayacaktır [2]. Bunun yanında YZ teknolojilerinin eğitim süreçlerinde giderek daha görünür hale gelmesi, öğretmen adaylarının yalnızca genel dijital yeterlikler bakımından değil, YZya ilişkin farkındalık ve okuryazarlık açısından da değerlendirilmesini gerekli kılmaktadır. Bu bağlamda öğretmen adaylarının dijital okuryazarlık düzeyleri ile YZ okuryazarlığı becerileri arasındaki ilişkinin ortaya konulması, öğretmen yetiştirme programlarının geliştirilmesi ve daha nitelikli öğrenme ortamlarının tasarlanması açısından önemli veriler sağlayabilir. Ayrıca bu tür çalışmalar, öğrencilerin bilimsel ve dijital yeterliklerini destekleyecek öğretim uygulamalarının güçlendirilmesine de katkı sunabilir [4].

G. Problem Durumu

Günümüzde eğitim sistemlerinin karşı karşıya olduğu temel sorunlardan biri, öğretmen adaylarının dijital çağın gerektirdiği yeterlikleri istenen düzeyde geliştirememesidir. Özellikle dijital okuryazarlık, bireyin dijital ortamlara erişebilmesi, bu ortamlarda sunulan bilgileri anlamlandırabilmesi ve söz konusu bilgileri eleştirel bir bakış açısıyla değerlendirebilmesi açısından temel bir beceri alanı olarak kabul edilmektedir [1]. Bu beceriler yalnızca dijital araçların kullanımını değil, aynı zamanda dijital içeriklerin bilinçli, sorgulayıcı ve amaç odaklı biçimde kullanılmasını da kapsamaktadır [10]. Öğretmen adaylarının dijital okuryazarlık düzeylerinin yetersiz kalması, onların gelecekteki öğretim uygulamalarında öğrencilerine dijital okuryazarlık, YZ okuryazarlığı ve bilgi okuryazarlığı gibi çağdaş becerileri kazandırmalarını güçleştirebilir. Böyle bir durum, öğrencilerin dijital çevrelerde karşılaştıkları bilgileri doğru değerlendirme, teknolojiyi verimli kullanma ve dijital öğrenme süreçlerine etkili katılım sağlama becerilerinin sınırlı kalmasına neden olabilir. Bunun yanı sıra YZ teknolojilerinin eğitim ortamlarında giderek daha yaygın biçimde kullanılmaya başlanması, öğretmen adaylarının bu teknolojilere ilişkin bilgi, farkındalık ve kullanım yeterliklerinin de incelenmesini gerekli hale getirmektedir [19]. Bu çerçevede araştırmada, öğretmen adaylarının dijital okuryazarlık düzeyleri ile YZ okuryazarlığı becerileri arasındaki ilişkinin incelenmesi amaçlanmaktadır. Elde edilecek bulguların, öğretmen eğitiminde dijital ve YZ temelli yeterliklerin güçlendirilmesine yönelik düzenlemelere katkı sağlaması beklenmektedir. Aynı zamanda bu araştırmanın, öğretim programlarının geliştirilmesi ve öğretmen adaylarının çağın gerektirdiği becerilerle donatılması açısından yol gösterici olacağı düşünülmektedir.

III. YÖNTEM

A. Araştırma Deseni

Bu araştırma, öğretmen adaylarının dijital okuryazarlık düzeyleri ile YZ okuryazarlığı becerileri arasındaki ilişkiyi ortaya koymak amacıyla nicel araştırma yaklaşımı çerçevesinde yürütülmüştür. Nicel araştırmalar, verilerin sayısal biçimde toplanmasına ve bu verilerin istatistiksel teknikler aracılığıyla analiz edilmesine dayanan bir araştırma sürecini ifade etmektedir [20]. Araştırmanın amacı doğrultusunda, katılımcılara ilişkin demografik özelliklerin belirlenmesinde betimsel istatistiklerden yararlanılmış; değişkenler arasındaki ilişkilerin incelenmesinde ise korelasyon ve regresyon analizleri kullanılmıştır. Bunun yanında cinsiyet ve sınıf düzeyi gibi demografik değişkenler açısından puan farklılıklarının belirlenmesine yönelik karşılaştırmalı istatistiksel işlemlere de yer verilmiştir.

B. Evren ve Örneklem

Araştırmanın evrenini Türkiye'de öğrenim gören öğretmen adayları oluşturmaktadır. Araştırmanın örneklemi ise Batı Karadeniz Bölgesi'nde yer alan bir devlet üniversitesinin eğitim fakültesinde öğrenim görmekte olan 214 öğretmen adayından meydana gelmektedir. Katılımcılar, araştırmaya dahil edilme sürecinde belirlenen ölçütler doğrultusunda seçilmiş ve çalışma bu grup üzerinden yürütülmüştür. Örneklem grubunun belirlenmesinde tercih edilen yaklaşım, araştırmanın uygulanabilirliğini sağlamak ve veri toplama sürecini yönetilebilir hale getirmek açısından önem taşımaktadır. Örnekleme sürecinde kullanılan yöntem, çalışmanın kapsamı ve erişilebilirlik koşulları dikkate alınarak belirlenmiştir. Alanyazında benzer örneklem gruplarıyla yürütülen çok sayıda araştırma bulunduğu görülmektedir [21]. Çalışma grubuna ilişkin demografik özellikler Tablo 1'de sunulmuştur.

TABLE I. ÇALIŞMA GRUBUNA AIT DEMOGRAFIK ÖZELLIKLER

Değişken	Alt Değişken	f	%
Cinsiyet	Kadın	148	69.16
	Erkek	66	30.84
Sınıf Düzeyi	1.sınıf	48	22.43

2.sınıf	51	23.83
3.sınıf	56	26.17
4.sınıf	59	27.57
Toplam	**214**	**100**

Tablo 1 incelendiğinde araştırmaya 214 öğretmen adayı katıldığı ve bu katılımcıların yüzde dağılımı ise %69.16 kadın (n=148) ve %30.84 (n=66) erkek şeklinde olduğu görülmektedir. Sınıf düzeyi değişkeni incelendiğinde 1.sınıfta bulunanların %22.43 (n=48), 2.sınıfta bulunanların %23.83 (n=51), 3.sınıfta bulunanların %26.17 (n=56) ve 4.sınıfta bulunanların ise %27.57 (n=59) olduğu görülmektedir.

C. Veri Toplama Araçları

Araştırmada veri toplama aracı olarak kişisel bilgi formu ile birlikte iki ayrı ölçek kullanılmıştır. Bu ölçeklerden ilki, [1] tarafından geliştirilen ve [2] tarafından Türkçeye uyarlanan **Dijital Okuryazarlık Ölçeği**dir. İkincisi ise [3] tarafından geliştirilen ve [4] tarafından Türkçeye uyarlanan **YZ Okuryazarlığı Ölçeği**dir. Dijital Okuryazarlık Ölçeği, katılımcıların dijital çevrelerde bilgiye ulaşma, bilgiyi anlama, değerlendirme ve uygulama becerilerini ölçmeye odaklanmaktadır. YZ Okuryazarlığı Ölçeği ise katılımcıların YZ ile ilgili bilgi düzeylerini, farkındalıklarını ve bu teknolojileri değerlendirme ile kullanma becerilerini belirlemeyi amaçlamaktadır. [1] tarafından geliştirilen ölçek, tek faktörde toplam 10 maddeden oluşmaktadır. YZ Okuryazarlığı Ölçeği ise öğrencilerin duygusal, davranışsal, bilişsel ve etik becerilerini değerlendirmek amacıyla tasarlanmış olan bir ölçektir. [3] tarafından geliştirilen bu ölçek beşli likert ve toplam 32 maddeden oluşmaktadır. Dijital ve YZ okuryazarlıklarını detaylı bir şekilde ölçülmekte ve bu ölçümler eğitim süreçlerinde önemli geri bildirimler sağlamaktadır.

D. Verilerin Analizi

Araştırma kapsamında elde edilen veriler, istatistiksel analizlerin yürütülmesi amacıyla SPSS paket programı kullanılarak çözümlenmiştir. Analiz sürecinin ilk aşamasında, öğretmen adaylarının dijital okuryazarlık ve YZ okuryazarlığı puanlarına ilişkin genel görünümü ortaya koymak amacıyla betimsel istatistiklerden yararlanılmıştır. Ardından söz konusu iki değişken arasındaki ilişkinin yönünü ve düzeyini belirlemek için korelasyon analizi gerçekleştirilmiştir. Son aşamada ise dijital okuryazarlık düzeyinin YZ okuryazarlığını ne ölçüde açıkladığını incelemek amacıyla regresyon analizine başvurulmuştur [22]. Bu doğrultuda, Dijital Okuryazarlık Ölçeğinden elde edilen puanların YZ Okuryazarlığı puanlarını anlamlı biçimde yordayıp yordamadığı değerlendirilmiştir.

E. Geçerlik ve Güvenirlik

Araştırmada kullanılan ölçme araçlarının geçerlik ve güvenirlik özellikleri, hem ölçeklerin geliştirilme ve uyarlama çalışmalarında rapor edilen bulgular hem de mevcut araştırmadan elde edilen iç tutarlılık katsayıları dikkate alınarak değerlendirilmiştir. Dijital Okuryazarlık Ölçeğinin özgün geliştirme ve Türkçeye uyarlama çalışmalarında, ölçme aracının geçerli ve güvenilir sonuçlar ürettiği belirtilmiştir. Benzer biçimde YZ Okuryazarlığı Ölçeği için de geliştirildiği ve uyarlandığı çalışmalarda yeterli düzeyde geçerlik ve güvenirlik değerleri rapor edilmiştir. Bunun yanında, bu araştırma kapsamında ölçeklerin iç tutarlılığını belirlemek amacıyla Cronbach's Alpha katsayıları hesaplanmıştır. Analiz sonucunda, Dijital Okuryazarlık Ölçeği için Cronbach's Alpha değeri 0.86, YZ Okuryazarlığı Ölçeği için ise 0.91 olarak bulunmuştur. Elde edilen bu katsayılar, kullanılan ölçeklerin yüksek düzeyde güvenilir olduğunu göstermektedir [23].

IV. BULGULAR

Araştırma kapsamında elde edilen veriler, öğretmen adaylarının dijital okuryazarlık düzeyleri ile YZ okuryazarlığı becerileri arasındaki ilişkiyi ortaya koymak amacıyla çeşitli istatistiksel analizler aracılığıyla incelenmiştir. Bu bölümde, öncelikle katılımcılara ilişkin betimsel bulgulara yer verilmiş; ardından cinsiyet ve sınıf düzeyi gibi değişkenlere göre yapılan karşılaştırmalı analiz sonuçları sunulmuştur. Ayrıca değişkenler arasındaki ilişkinin belirlenmesine yönelik korelasyon analizi ile yordama gücünü ortaya koymaya yönelik regresyon analizi bulgularına da yer verilmiştir.

A. Betimsel ***Analiz***

Öğretmen adaylarının dijital okuryazarlık ve YZ okuryazarlığı düzeylerini belirlemek amacıyla betimsel analiz yapılmıştır. Öğretmen adaylarının dijital okuryazarlık ve YZ okuryazarlığı puanlarının ortalama ve standart sapma değerlerine ilişkin betimsel istatistik analiz bulguları Tablo 2'de sunulmuştur.

TABLE II. BETIMSEL İSTATISTIK ANALIZI

Değişken	n	Ortalama	Standart Sapma
Dijital Okuryazarlık	214	3.87	0.46
YZ Okuryazarlığı	214	3.54	0.63

Tablo 2'de görüldüğü gibi, katılımcıların dijital okuryazarlık seviyelerinin yüksek, YZ okuryazarlığı seviyelerinin ise orta düzeyin üstünde olduğu tespit edilmiştir. Bu bulgular, öğretmen adaylarının dijital çevreye ulaşma, dijital bilgiyi anlama, pratik yapma, değerlendirme vb. becerilerin doğru orantılı olarak artmış olduğunu göstermektedir.

B. Cinsiyete Göre İstatistikler

Öğretmen adaylarının dijital okuryazarlık puanları ile YZ okuryazarlığı puanları arasında cinsiyet değişkenine ilişkin anlamlı bir farklılık oluşturup oluşturmadığı incelenmiştir. Bu bağlamda cinsiyete göre bağımsız örneklemler t-testi yapılması amacıyla gerekli istatistikler hesaplanmıştır. Tablo 3 ve Tablo 4'te sırasıyla dijital okuryazarlık ve YZ okuryazarlığı değişkenleri ile ilgili bağımsız örneklemler t-testi değerleri bulunmaktadır.

TABLE III. CINSIYETE GÖRE DIJITAL OKURYAZARLIK IÇIN BAĞIMSIZ ÖRNEKLEMLER T-TESTI

Grup	n	Ort.	Ss	t	sd	p
Kadın	148	3.57	0.58	-1.39	212	0.166
Erkek	66	3.69	0.62			

Tablo 3'de görüldüğü gibi, cinsiyet değişkeni sonuçları bağımsız gruplar t testi incelemesi yapıldığında, öğretmen adaylarının dijital okuryazarlık test puanları arasında cinsiyet değişkeni açısından anlamlı farklılık oluşturmadığı [t(212)= -1.39; p=0.166>0.05] belirlenmiştir.

TABLE IV. CINSIYETE GÖRE YZ OKURYAZARLIĞI IÇIN BAĞIMSIZ ÖRNEKLEMLER T-TESTI

Grup	n	Ort.	Ss	t	sd	p
Kadın	148	3.49	0.61	-1.72	212	0.087
Erkek	66	3.65	0.66			

Tablo 4'te görüldüğü gibi, cinsiyet değişkeni sonuçları bağımsız gruplar t testi açısından incelendiğinde, öğretmen adaylarının YZ okuryazarlığı test puanları arasında cinsiyet değişkeni açısından anlamlı farklılık oluşturmadığı [t(212)= -1.72; p=0.087>0.05] belirlenmiştir.

C. Sınıf Düzeyine Göre İstatistikler

Öğretmen adaylarının dijital okuryazarlık ve YZ okuryazarlığı test puanları arasında sınıf düzeyi değişkeni açısından anlamlı farklılık oluşturup oluşturmadığı incelenmiştir. Bu bağlamda sınıf düzeyine göre ANOVA testi yapılması amacıyla gerekli istatistikler hesaplanmıştır. Tablo 5 ve Tablo 6'da sınıf düzeyine göre dijital okuryazarlık ve YZ okuryazarlığına ilişkin ANOVA testi değerleri bulunmaktadır.

TABLE V. SINIF DÜZEYINE GÖRE DIJITAL OKURYAZARLIK PUANLARI IÇIN ANOVA TESTI

Grup	n	Ort.	Ss	F	p
1. Sınıf	48	3.48	.59	3.12	0.027
2. Sınıf	51	3.55	.61		
3. Sınıf	56	3.67	.56		
4. Sınıf	59	3.73	.60		

Tablo 5'te görüldüğü gibi, sınıf düzeyi değişkeni sonuçlarının tek yönlü varyans analizi açısından incelemesi yapıldığında, katılımcıların dijital okuryazarlığa ilişkin test puanları arasında, sınıf düzeyi değişken bakımından anlamlı farklılık bulunduğu [F(3.210)= 3.12; p=0.027<0.05] tespit edilmiştir. Dijital okuryazarlık puanları arasındaki bu farklılığın üst sınıflar lehine olduğu söylenebilir.

TABLE VI. SINIF DÜZEYINE GÖRE YZ OKURYAZARLIĞI PUANLARI IÇIN ANOVA TESTI

Grup	n	Ort.	Ss	F	p
1. Sınıf	48	3.42	.63	2.84	0.039
2. Sınıf	51	3.50	.64		
3. Sınıf	56	3.58	.60		
4. Sınıf	59	3.68	.62		

Tablo 6'da görüldüğü gibi, sınıf düzeyi değişkeni sonuçlarının tek yönlü varyans analizi açısından incelemesi yapıldığında, öğretmen adaylarının YZ okuryazarlığı test puanları arasında sınıf düzeyi değişkeni açısından anlamlı farklılık bulunduğu [F(3.210)= 2.84; p=0.039<0.05] belirlenmiştir. YZ okuryazarlığı puanları arasındaki bu farklılığın da üst sınıflar lehine olduğu söylenebilir.

D. YZ deneyimine Göre İstatistikler

Öğretmen adaylarının dijital okuryazarlık ve YZ okuryazarlığı ANOVA testi puanları arasında YZ deneyimi değişken bakımından anlamlı farklılık oluşturup oluşturmadığı incelenmiştir. Bu bağlamda YZ deneyimine göre ANOVA testi yapılması amacıyla gerekli istatistikler hesaplanmıştır. YZ okuryazarlığı deneyimine göre dijital okuryazarlık ve YZ okuryazarlığına ilişkin ANOVA testi değerleri sırasıyla Tablo 7 ve Tablo 8'de bulunmaktadır.

TABLE VII. YZ DENEYIMI DÜZEYINE GÖRE DIJITAL OKURYAZARLIK PUANLARI IÇIN ANOVA TESTI

Grup	n	Ort.	Ss	F	p
Düşük	62	3.39	.57	8.47	0.000
Orta	88	3.61	.55		
Yüksek	64	3.82	.58		

Tablo 7'de görüldüğü gibi, YZ deneyimi değişkeni sonuçlarının ANOVA testi açısından incelemesi yapıldığında, öğretmen adaylarının dijital okuryazarlık test puanları arasında YZ deneyimi değişkeni bakımından anlamlı farklılık bulunduğu [F(2.211)= 8.47; p=0.000<0.05] belirlenmiştir. Bu farklılığın YZ deneyim düzeyi yüksek olan öğretmen adayları lehine olduğu görülmektedir.

TABLE VIII. YZ DENEYIMI DÜZEYINE GÖRE YZ OKURYAZARLIĞI PUANLARI IÇIN ANOVA TESTI

Grup	n	Ort.	Ss	F	p
Düşük	62	3.21	.58	21.36	0.000
Orta	88	3.54	.56		
Yüksek	64	3.93	.51		

Tablo 8'de görüldüğü gibi, YZ deneyimi değişkeni sonuçlarının ANAOVA testi açısından incelemesi yapıldığında, öğretmen adaylarının YZ okuryazarlığı test puanları arasında YZ deneyimi değişkeni bakımından anlamlı farklılık bulunduğu [F(2.211)= 21.36; p=0.000<0.05] belirlenmiştir. Bu farklılığın YZ deneyim düzeyi yüksek olan öğretmen adayları lehine olduğu söylenebilir.

Öğretmen adaylarının dijital okuryazarlık ve YZ okuryazarlığı puanları cinsiyet, sınıf düzeyi ve YZ deneyimi değişkenleri açısından incelenmiştir. Cinsiyet değişkeni ile ilgili yapılan t testi sonuçları hem dijital okuryazarlık hem de YZ okuryazarlığı puanlarında anlamlı farklılık olmadığını göstermiştir. Buna karşılık sınıf düzeyi ve YZ deneyimi değişkenleri açısından yapılan ANOVA testi sonuçları, özellikle üst sınıflarda öğrenim gören ve YZ deneyimi daha

yüksek olan öğretmen adaylarının, doğru orantılı olarak daha yüksek puanlara sahip olduğu görülmektedir.

E. Korelasyon Analizi

Dijital okuryazarlık ile YZ okuryazarlığı becerileri arasındaki ilişkiyi belirlemek amacıyla Pearson korelasyon analizi yapılmıştır. Söz konusu korelasyon analiz bulguları Tablo 9'da görülmektedir.

TABLE IX. KORELASYON ANALIZI

Değişken Çifti	Korelasyon Katsayısı (r)	*p*
Dijital Okuryazarlık- YZ Okuryazarlığı	0.56	0.01

Tablo 9'da görülen korelasyon analizi sonuçlarına göre, dijital okuryazarlık ile YZ okuryazarlığı becerileri arasında pozitif ve anlamlı bir ilişki bulunmaktadır ($r=0.56$, $p<0.01$). Bu sonuç, dijital okuryazarlık düzeyi yüksek olan öğretmen adaylarının aynı zamanda YZ okuryazarlığı becerilerinin de yüksek olduğunu göstermektedir. Bu durum hem dijital bilgiyi hem de YZ öğrenme ortamlarını kullanabilme becerileri arasında bir ilişki olduğunu göstermektedir.

A. Regresyon Analizi

Öğretmen adaylarının dijital okuryazarlık düzeylerinin, YZ okuryazarlığı düzeylerini ne derece etkilediğini tespit etmek için regresyon analizi uygulanmaktadır. Dijital okuryazarlık düzeyi bağımsız değişken, YZ okuryazarlığı düzeyi ise bağımlı değişken olarak belirlenmiş ve bu değişkenler arasında regresyon analizi yapılmıştır. Basit doğrusal regresyon için oluşturulan örnek tabloya göre dijital okuryazarlık toplam puanı, YZ okuryazarlığı toplam puanını anlamlı biçimde yordamaktadır.

TABLE X. REGRESYON ANALIZI

Yordayıcı	B	Std. Hata	β	t	p
Sabit	0.91	0.29	-	3.14	0.002
Dijital Okuryazarlık	0.37	0.07	0.34	5.29	<0.01

Tablo 10'da görüldüğü gibi yapılan regresyon analiz sonucu, dijital okuryazarlık düzeyinin YZ okuryazarlığı becerisi üzerinde anlamlı ve pozitif bir etkisi olduğunu göstermektedir ($B=0.37$, $\beta=0.34$, $p<0.01$). Bu sonuca göre, dijital okuryazarlık düzeyinin artması halinde YZ okuryazarlığı becerisinin de paralel olarak artacağı görülmektedir. Dolayısıyla regresyon analizi, katılımcıların dijital okuryazarlık düzeyleri yüksek ise YZ okuryazarlığı becerilerinin de doğru orantılı bir şekilde yüksek olacağını göstermektedir.

V. SONUÇ VE ÖNERİLER

Bu araştırma, öğretmen adaylarının dijital okuryazarlık ve YZ okuryazarlığı düzeyleri arasında meydana gelen ilişkiyi incelemek için yapılmıştır. Elde edilen sonuçlara göre, öğretmen adaylarının dijital okuryazarlık düzeyleri genel olarak yüksek olduğu görülmektedir. Benzer şekilde YZ okuryazarlığı becerileri de ortalamanın üzerinde çıkmıştır. Korelasyon analizi ve regresyon analizi sonuçları, dijital ve YZ okuryazarlıklarına ilişkin beceriler arasında anlamlı ve pozitif bir ilişki olduğunu göstermektedir. Bu durumda öğretmen adaylarının dijital çevreye ulaşma ve YZ uygulamalarını kullanma kabiliyetlerinin gelişmiş olduğu görülmektedir.

Bu çalışma, öğretmen adaylarının dijital okuryazarlık ile YZ okuryazarlığı düzeyleri arasında, eğitimde dijital dönüşüm bağlamında bir ilişki olabileceğine ve bunun önemli bir işaret olduğunu göstermektedir.

Özellikle dijital okuryazarlığın artması, öğretmen adaylarının YZ araçlarını tanıma, değerlendirme ve eğitim amaçlı kullanma kapasitesini güçlendirebilir. Bilişsel boyutun da anlamlı bir katkı sağlaması önemlidir. Çünkü YZ araçlarının eğitimde işlevsel olabilmesi, yalnızca komut verme becerisine değil; üretilen metni sorgulama, alternatifleri karşılaştırma, yanlılık ve hata olasılıklarını değerlendirme gibi üst düzey becerilere bağlıdır. Bu yönüyle dijital okuryazarlık, öğretmen adaylarının YZ çıktılarını pedagojik filtrelerden geçirerek kullanabilmesinin ön koşulu gibi görünmektedir.

[12] ve [14] tarafından yapılan çalışmalardaki öğretmen eğitiminde üretken YZ ile ilgili değerlendirmeleri, öğretmen yetiştirme programlarının YZ/Üretken YZ boyutunu açık şekilde içermesi gerektiğini göstermektedir.

[15] tarafından yapılan çalışma, teknik dijital okuryazarlığın artmasının YZ kaygısını azalttığını göstermektedir. Bu sonuç, YZ ile çalışma deneyimi düşük olan öğretmen adaylarında görülebilecek çekingenlik ya da mesleki belirsizlik duygularının, iyi yapılandırılmış dijital okuryazarlık eğitimleriyle azaltılabileceğini düşündürmektedir.

[16] ve [24] tarafından yapılan çalışma, dijital okuryazarlığın 21. yüzyıl öğretim yeterlikleriyle ilişkisine dair bulguları da bu çalışmanın gerekçesini desteklemektedir. Öğretmen adayları dijital çevreyi [25] ve bu çevre kapsamındaki dijital bilgiyi ne kadar iyi bilir ve yönetirse; teknoloji, pedagoji ve alan bilgisini bütünleştiren uygulamalar üretme olasılıkları o kadar yükselmektedir. Bu nedenle öğretmen eğitiminde dijital okuryazarlık ile YZ kullanımının birbirinden bağımsız iki alan olarak değil, birbirini güçlendiren bir yeterlik kümesi olarak ele alınması gerekir.

Araştırma çerçevesi ve temsili bulgular birlikte değerlendirildiğinde, öğretmen adaylarının dijital okuryazarlık düzeyleri yükseldikçe YZ destekli öğrenme ortamlarını daha bilinçli ve daha sık kullanma eğiliminde oldukları söylenebilir. Öğretmen yetiştirme programlarında 'üretken YZ ve öğretim tasarımı', 'dijital içerik doğrulama', 'YZ etiği ve veri güvenliği' ve 'alan öğretiminde YZ uygulamaları' gibi modüllerin zorunlu/seçmeli ders düzeyinde yapılandırılması önerilebilir. Öğretmen adaylarının YZ destekli deney tasarımı, kavram yanılgısı analizi, çoklu temsil üretimi ve öğrenci düzeyine uygun açıklama hazırlama gibi uygulamalarla desteklenmesi yararlı olabilir.

İzleyen çalışmalarda nicel verilerin nitel görüşmelerle desteklenmesi, öğretmen adaylarının YZ kullanımını yalnızca sıklık temelinde değil; kalite, etik farkındalık ve pedagojik derinlik temelinde de incelemeyi mümkün kılacaktır.

Araştırma çerçevesi ve temsili bulgular birlikte değerlendirildiğinde, öğretmen adaylarının dijital okuryazarlık düzeyleri yükseldikçe YZ okuryazarlığı puanlarının da artabileceği öngörülmektedir. Bu nedenle öğretmen yetiştirme programlarında hem dijital okuryazarlığa hem de YZ okuryazarlığına yönelik sistematik öğrenme yaşantıları planlanmalıdır.

Eğitimde dijital dönüşüm bağlamında buradan şu sonuç çıkarılabilir; Öğretmen adaylarına yalnızca yeni araçlar tanıtmak yeterli değildir. Bunun yerine öğretim tasarımında YZ okuryazarlığı, etik sınırlar, akademik dürüstlük, veri gizliliği, kaynak doğrulama, önyargı farkındalığı ve öğrenci öğrenmesini destekleyen özgün senaryo tasarımları birlikte öğretilmelidir.

REFERENCES

[1] Ng, W. (2012). Can we teach digital natives digital literacy? Computers & Education, 59(3), 1065-1078.

[2] Üstündağ, M. T., Güneş, E., & Bahçivan, E. (2017). Dijital Okuryazarlık Ölçeğinin Türkçeye Uyarlanması ve Fen Bilgisi Öğretmen Adaylarının Dijital Okuryazarlık Durumları. Journal of Education and Future, 12, 19-29. https://izlik.org/JA62NF53DZ

[3] Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. British Journal of Educational Technology, 55(3), 1082–1104. https://doi.org/10.1111/bjet.13411

[4] Gökoğlu, S. Erdoğdu, F., & Altun, E. (2025). Artificial intelligence literacy scale: Turkish adaptation study. Anadolu Journal of Educational Sciences International, 15(4), 1327-1357. https://doi.org/10.18039/ajesi.1699897

[5] UNESCO. (2025). AI and education: Guidance for policy-makers. UNESCO.

[6] UNESCO. (2026). Artificial intelligence in education. UNESCO Digital Education.

[7] OECD. (2021). OECD digital education outlook 2021: Pushing the frontiers with artificial intelligence, blockchain and robots. OECD Publishing.

[8] OECD. (2025). Preparing teachers for digital education. OECD Publishing.

[9] Yanarateş, E. (2021). Fen bilimleri öğretimi ve teknoloji kullanımında güncel yaklaşımlar. T. Talan (Ed.), Eğitimde Dijitalleşme ve Yeni Yaklaşımlar içinde (s. 59-104), Efe Akademi Yayınları

[10] Falloon, G. (2020). From digital literacy to digital competence: The teacher digital competency (TDC) framework. Educational Technology Research and Development, 68(5), 2449-2472. https://doi.org/10.1007/s11423-020-09767-4

[11] Çelebi, C., Yilmaz, F., Demir, U. ve Karakuş, F. (2023). Yapay zekâ okuryazarlığı: Bir ölçek uyarlama çalışması. Instructional Technology and Lifelong Learning, 4(2), 291–306. https://doi.org/10.52911/itall.1401740

[12] Prilop, C. N., Weber, S., & Kleinknecht, M. (2025). Generative AI in teacher education: Educators' perceptions of transformative potentials and the triadic nature of AI literacy explored through AI-enhanced methods. Computers and Education Open, 7.

[13] Spante, M., Hashemi, S. S., Lundin, M., & Algers, A. (2018). Digital competence and digital literacy in higher education research: Systematic review of concept use. Cogent Education, 5(1), 1519143.

[14] Ayanwale, M. A., Sanusi, I. T., Aruleba, K., Oyelere, S. S., Haruna, H., & Agbo, F. J. (2024). Examining artificial intelligence literacy among pre-service teachers for future classrooms. Computers and Education: Artificial Intelligence, 6, 100229.

[15] Ayduğ, D. (2025). Are Turkish pre-service teachers worried about AI? A study on AI anxiety and digital literacy. AI & Society. Advance online publication.

[16] Aslan, S., Alanoğlu, M., & Karabatak, S. (2025). Enhancing 21st-century teaching competencies: The key role of digital literacy in connecting pre-service teachers' TPACK. Active Learning in Higher Education. Advance online publication.

[17] Long, D., & Magerko, B. (2020, April 25–30). What is AI literacy? Competencies and design considerations [Conference presentation]. CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA. https://doi.org/10.1145/3313831.3376727

[18] Kandlhofer, M., Steinbauer, G., Hirschmugl-Gaisch, S., & Huber, P. (2016, October 12–15). Artificial intelligence and computer science in education: From kindergarten to university [Conference presentation]. IEEE Frontiers in Education Conference, Erie, PA, USA. https://doi.org/10.1109/FIE.2016.7757570

[19] Akyürek, M. İ. (2025). Yapay zekâ Okuryazarlığı Ölçeğinin Türkçe Uyarlaması: Geçerlik ve Güvenirlik Çalışması (Eğitim Paydaşlarına Yönelik). Gazi Eğitim Fakültesi Dergisi, 45(2), 649-663. https://doi.org/10.17152/gefad.1469370

[20] Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches. Thousand Oaks, CA: SAGE Publications.

[21] Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. American Journal of Theoretical and Applied Statistics, 5(1), 1-4.

[22] Field, A. (2017). Discovering statistics using IBM SPSS statistics. Thousand Oaks, CA: SAGE Publications.

[23] Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. International Journal of Medical Education, 2, 53-55.

[24] Kaya, F., Aydin, F., Schepman, A., Rodway, P., Yetişensoy, O., & Kaya, M. D. (2022). The roles of personality traits, ai anxiety, and demographic factors in attitudes toward artificial intelligence, Int. J. of Human–Computer Interaction (40), 497-514, Doi: 10.1080/10447318.2022.2151730

[25] Yanarateş, E. (2022). Dijital çevre ve dijital atık. M. Baş, İ. Erdoğan-Tarakçı, R. Aslan (Eds.). Dijitalleşme içinde (ss.159-176). Efe Akademi Yayınları.

Phishing Email Detection Using a Hybrid Model Integrating Transformer Architectures and Social Engineering Analysis

Nada Nadim Al Jondi
Department of Computer Engineering
Karabuk University
Karabuk 78050, TURKIYE
2328126454@ogrenci.karabuk.edu.tr

Oğuzhan Menemencioğlu
Department of Computer Engineering
Karabuk University
Karabuk 78050, TURKIYE
omenemencioglu@karabuk.edu.tr

***Abstract*—Phishing remains a significant cybersecurity threat due to its reliance on social engineering tactics that exploit human psychological and behavioral vulnerabilities. While transformer-based architectures like BERT and RoBERTa have improved detection by representing contextual features, many current deep learning models treat phishing detection as a generic text classification task. This research addresses the deficiency in previous models that lack structured behavioral analysis and interpretability. We propose a hybrid model that integrates the DistilRoBERTa architecture with behavioral indicators derived from the V-Triad framework, which categorizes attacks across dimensions of credibility, contextual compatibility, and urgency. Our methodology employs a late-fusion strategy, combining deep contextual logits with binary behavioral features and processing them with a Logistic Regression classifier to maintain transparency. Furthermore, an explanatory mechanism is incorporated to align model decisions with observable textual evidence. Experimental evaluations on a standardized dataset demonstrate that the proposed model achieves high performance, reaching an overall accuracy of 98.45 percent and a recall of 99.4 percent. These results suggest that fusing deep linguistic representations with structured behavioral grounding provides a robust and explainable solution for identifying sophisticated phishing attempts.**

Keywords—phishing email, DistilRoBERTa, social engineering, V-Triad

I. INTRODUCTION

Artificial Intelligence (AI) is formally defined as the branch of computer science dedicated to developing computational systems capable of performing tasks requiring human cognitive abilities, such as reasoning, learning, and autonomous decision-making [1], [2]. Within this domain, Natural Language Processing (NLP) enables machines to interpret human language in a contextually and semantically coherent manner. The evolution of NLP has been driven by transformer-based architectures, marking a paradigm shift toward deeper contextual understanding beyond surface-level lexical features [3], [4]]. This transformation has critical implications for cybersecurity, specifically in the detection of phishing attacks [5]. Phishing attacks are sophisticated forms of social engineering characterized by psychological manipulation, stylistic imitation, and contextual personalization aimed at extracting sensitive information [6]. Although transformer-based models have demonstrated a superior ability to identify malicious linguistic patterns compared to traditional methods, current research often suffers from a lack of interpretability and grounding in behavioral frameworks [7]. These models typically function as "black boxes," providing high quantitative accuracy without explaining the psychological tactics used by attackers. This research addresses these deficiencies by combining deep contextual linguistic modeling with a structured psychological-behavioral framework.

A. Transformer-Based Methods

Literature indicates a growing trend toward using transformer-based models for identifying phishing attacks due to their superior contextual understanding. Comparative studies between BERT and RoBERTa reveal that RoBERTa consistently outperforms BERT, achieving better accuracy and robustness, particularly in adversarial environments where phishing emails utilize obfuscation strategies [7]. However, BERT is often constrained by a 512-token input limit, which can negatively impact performance compared to architectures like GloVe-CNN in specific contexts[8]. Other optimizations involve using architectures like DistilRoBERTa for computational efficiency or balancing datasets via ADASYN, which can yield accuracies up to 99.00 percent, though these remain focused on general classification rather than phishing-specific behavioral analysis[9].

B. Hybrid and Explainable Approaches

To address transparency, recent work has integrated transformers with explainability tools. The combination of transformers and BiLSTMs has shown excellent results but is susceptible to overfitting on small or synthetic datasets [10]. The LITA framework represents a significant advancement by integrating RoBERTa with LIME and Transformer-Interpret to provide token-level explanations, achieving 98.45 percent accuracy and near-perfect recall [11].

C. Dataset-Specific Comparisons

The SpamAssassin corpus is frequently utilized to evaluate models. Traditional machine learning approaches, including Random Forest and XGBoost, provide high accuracy on this dataset but rely on surface representations like TF-IDF [12]. Evaluations of corporate-level tools such as Rspamd, SpamAssassin, and Dspam highlight that while

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

effective for basic filtering, they may not capture the nuances of modern social engineering [13]. Hybrid models combining CNNs and XGBoost have achieved 98 percent accuracy but often ignore behavioral interpretation [14].

II. Background

Transformer architectures replace recurrence with self-attention mechanisms, enabling parallel processing of tokens and effective modeling of long-range dependencies [15] [16]. While BERT uses bidirectional contextualization, RoBERTa optimizes this by removing next-sentence prediction tasks. DistilRoBERTa further reduces computational demands through knowledge distillation without sacrificing contextualization capabilities [15],[17].The V-Triad framework provides a behavioral model for analyzing phishing emails based on three dimensions:

- Credibility: Established through institutional imitation and authoritative language [18]
- Contextual Compatibility: Aligning the message with the user's expected activities [19].
- Urgency: Inducing rapid decision-making through time pressure, which limits analytical processing [20].

III. Methodology

A. Dataset and Preprocessing

This study utilizes a binary-labeled dataset derived from the SpamAssassin corpus, consisting of 5,809 messages (4,091 legitimate and 1,718 phishing). Subject and body were merged to preserve semantic cues. A stratified 80/20 train-test split was used, and RandomOverSampler was applied exclusively to the training data to address class imbalance without inducing data leakage.

B. Feature Extraction and V-Triad Integration

The framework extracts manual behavioral features based on the V-Triad model. To ensure technical rigor, the model distinguishes between the behavioral dimension and the linguistic feature:

- Credibility: Terms related to institutions or security (e.g., "bank," "login," "verify").
- Compatibility: Patterns reflecting formal or user-directed communication (e.g., "dear," "customer").
- Consistency: This feature represents the linguistic expressions indicating the Urgency dimension (e.g., "urgent," "immediately," "act now").These are extracted as binary features indicating the presence or absence of specific patterns.

C. Technical Pipeline and Classification

The pipeline follows a late fusion strategy. The input email is processed in parallel:

- Contextual Branch: DistilRoBERTa (distilroberta-base) is fine-tuned with max_length=128 and num_train_epochs=2 to extract deep contextual logits.
- Behavioral Branch: The V-Triad extraction function identifies the binary features (Credibility, Compatibility, Consistency).The logits and binary features are concatenated into a unified vector. Because the vector contains mixed scales (continuous logits and binary behavioral indicators), StandardScaler is applied to ensure the classifier treats features with equal weight. The final classification is performed by a Logistic Regression model, which computes class probabilities via a sigmoid function.

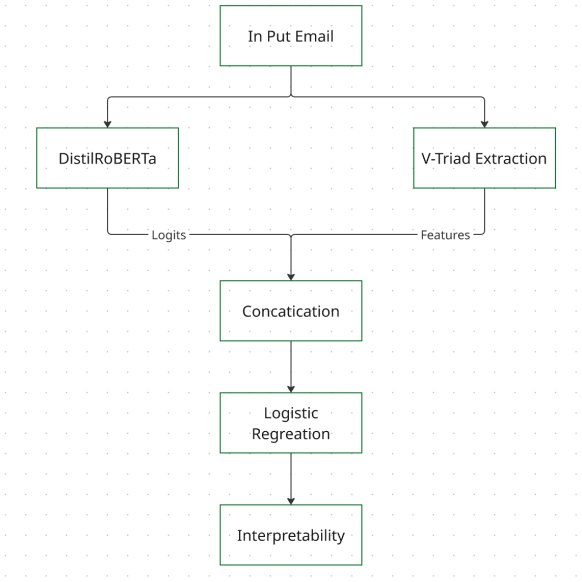


Fig. 1. illustrating the method of the model

D. Interpretability Mechanism

After classification, an explanatory stage uses a prompt template to provide rationale for the decision. For instance, if the Logistic Regression model predicts a phishing probability of 0.88, the explainer highlights specific evidence such as "Probability is 0.71 for the phishing class" and "Contains 'Urgent Action' keyword." This aligns the output directly with the V-Triad dimensions used during extraction.

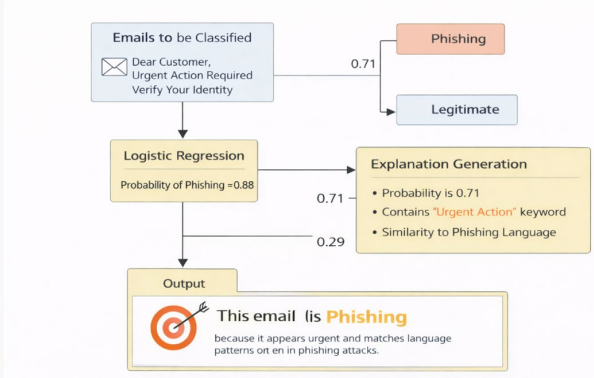


Fig. 2. Generated output

IV. Results and Discussion

The section presents the results of the approach and discussion.

A. Performance Metrics

The hybrid framework demonstrated high effectiveness in retrieving security threats while maintaining low false-positive rates.

TABLE I. PERFORMANCE METRICS OF THE HYBRID MODEL

Metric	Value	Interpretation
Accuracy	0.9845	High overall classification correctness
Precision	0.9552	Low false positive rate
Recall	0.9942	Strong phishing detection capability
F1-score	0.9743	Balanced precision-recall trade-off

[a.] Sample of a Table footnote. (*Table footnote*)

B. Class-Wise Analysis

For Class 0 (Legitimate), the model achieved 1.00 precision and 0.98 recall. For Class 1 (Phishing), it reached 0.96 precision and a significant 0.99 recall. The high recall for phishing confirms the model's sensitivity to fraudulent indicators even when they mimic legitimate communication styles.

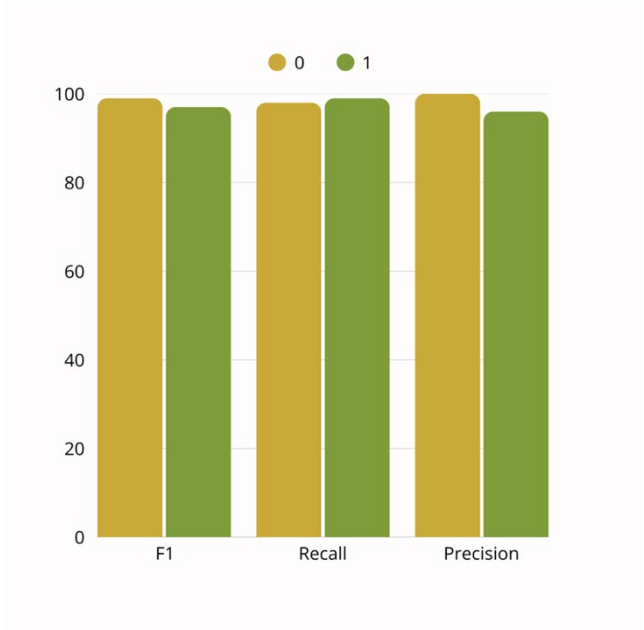


Fig. 3. Class-wise performance

C. Comparative Discussion

The proposed model ranks among the highest-performing solutions compared to studies using the SpamAssassin dataset. Notably, it outperforms standard ML models that rely on surface representations 14 and existing filtering tools 15.

TABLE II. COMPARISON WITH SPAMASSASSIN-BASED STUDIES

Study	Method	SpamAssassin Result
[12]	TF-IDF + Random Forest, Logistic Regression, Naive Bayes, XGBoost	Accuracy = **0.98,** F1-score =**0.96**
[13]	Rspamd, SpamAssassin, Dspam	Precision = **0.805**, Recall = **0.934**, F1 = **0.865**. **Dspam:** Precision = **0.952**, Recall = **0.990**, F1 = **0.971**
[14]	CNN + DFT + XGBoost	**Accuracy = 98%**

V. CONCLUSION

This study demonstrates the potential of integrating deep linguistic representations from DistilRoBERTa with structural behavioral features from the V-Triad framework. By combining semantic analysis with explicit indicators of credibility, compatibility, and urgency, the hybrid model achieves a superior balance between precision and recall. Logistic Regression serves as a sound methodological basis for interpretability, allowing for transparent behavioral analysis of phishing predictions without sacrificing computational performance.

ACKNOWLEDGMENT *(Heading 5)*

This paper summarizes a portion of Nada Nadim Al Jondi M.Sc. thesis research at Karabuk University in 2026, supervised by Dr. Oğuzhan Menemencioğlu.

REFERENCES

[1] R. Yousri and G. Hany, "Artificial Intelligence: An overview."

[2] N. Kühl, M. Schemmer, M. Goutier, and G. Satzger, "Artificial intelligence and machine learning," *Electronic Markets*, vol. 32, no. 4, pp. 2235–2244, Dec. 2022, doi: 10.1007/s12525-022-00598-0.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," May 2019, [Online]. Available: http://arxiv.org/abs/1810.04805

[4] C. E. Amaechi and O. C Okeke, "Leveraging NLP and Deep Learning for Phishing Detection and Anti-Phishing Training in Nigeria: A Focus on Localized Tactics and Cultural Factors," *International Research Journal of Innovations in Engineering and Technology*, vol. 08, no. 10, pp. 01–06, 2024, doi: 10.47001/irjiet/2024.810001.

[5] B. Gogoi and T. Ahmed, "Phishing and Fraudulent Email Detection through Transfer Learning using pretrained transformer models," in *INDICON 2022 - 2022 IEEE 19th India Council International Conference*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/INDICON56171.2022.10040097.

[6] F. Heiding, B. Schneier, A. Vishwanath, J. Bernstein, and P. S. Park, "Devising and detecting phishing emails using large language models", doi: 10.1109/ACCESS.2017.DOI.

[7] A. Alhuzali, A. Alloqmani, M. Aljabri, and F. Alharbi, "In-Depth Analysis of Phishing Email Detection: Evaluating the Performance of Machine Learning and Deep Learning Models Across Multiple Datasets," *Applied Sciences (Switzerland)*, vol. 15, no. 6, Mar. 2025, doi: 10.3390/app15063396.

[8] S. Giri, S. Banerjee, K. Bag, and D. Maiti, "Comparative Study of Content-Based Phishing Email Detection Using Global Vector (GloVe) and Bidirectional Encoder Representation from Transformer (BERT) Word Embedding Models," in

2022 1st International Conference on Electrical, Electronics, Information and Communication Technologies, ICEEICT 2022, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICEEICT53079.2022.9768612.

[9] S. Jamal, H. Wimmer, and I. Sarker, "An Improved Transformer-based Model for Detecting Phishing, Spam, and Ham: A Large Language Model Approach," Dec. 01, 2023. doi: 10.21203/rs.3.rs-3608294/v1.

[10] S. Vasudevan and V. Govindan, "Context-Aware Phishing: Leveraging Transformer and BiLSTM Networks for Cybersecurity," in *Proceedings of the 9th International Conference on Inventive Systems and Control, ICISC 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 520–526. doi: 10.1109/ICISC65841.2025.11188330.

[11] M. A. Uddin, M. Mahiuddin, and I. H. Sarker, "An Explainable Transformer-based Model for Phishing Email Detection: A Large Language Model Approach," Aug. 2025, [Online]. Available: http://arxiv.org/abs/2402.13871

[12] S. Twum, "Evaluation of Machine Learning Techniques for Identifying Phishing Emails: A Case Study with the Spam Assassin Dataset.," *INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, vol. 09, no. 04, pp. 1–9, Apr. 2025, doi: 10.55041/ijsrem45149.

[13] O. Ivan, L. Stepan, and S. Vladyslav, "Analysis and testing of systems countering phishing and social engineering attacks at the corporate level," *Int. J. Inf. Secur.*, vol. 25, no. 2, Apr. 2026, doi: 10.1007/s10207-026-01238-w.

[14] A. A. Ali and A. A. Abdullah, "Email Spam Detection: A Novel Hybrid Approach Using Machine and Deep Learning Techniques," *International Journal of Intelligent Engineering and Systems*, vol. 18, no. 7, pp. 157–174, 2025, doi: 10.22266/ijies2025.0831.12.

[15] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI Open*, vol. 3, pp. 111–132, Jan. 2022, doi: 10.1016/j.aiopen.2022.10.001.

[16] X. Song *et al.*, "Transformer: A Survey and Application."

[17] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," Mar. 2020, [Online]. Available: http://arxiv.org/abs/1910.01108

[18] ROBERT B. CIALDINI PH.D., *NATIONAL BESTSELLERNFLUENGEThe Psychology of Persuasion*. 2006.

[19] A. Vishwanath, T. Herath, R. Chen, J. Wang, and H. R. Rao, "Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model," *Decis. Support Syst.*, vol. 51, no. 3, pp. 576–586, Jun. 2011, doi: 10.1016/j.dss.2011.03.002.

[20] N. Security Agency and T. Transfer Program, "Building a science of cybersecurity: The next move."

FarmEcho: A Blockchain-LLM Framework for QR-Triggered Multilingual Agricultural Traceability with On-Chain Audit Trail

Abhishek Raut
Computer Engineering
VPKBIET
Baramati, India
abhishekraut980@gmail.com

Arvind Jagtap
Computer Engineering
VPKBIET
Baramati, India
arvind.jagtap@vpkbiet.org

ABSTRACT

Agricultural supply chains suffer from data opacity and lack of consumer-accessible traceability. Existing blockchain systems ensure immutability but present records as static, machine-readable outputs inaccessible to ordinary consumers. This paper presents FarmEcho, a framework integrating Ethereum blockchain traceability with a QR-triggered conversational LLM interface. Supply chain data across four stages is registered via a Solidity smart contract ensuring cryptographic auditability. Scanning a QR code opens a conversational AI interface enabling consumers to query food safety information in English, Hindi, or Marathi. An on-chain audit trail detects modifications and identifies responsible parties by wallet address. FarmEcho is the first framework to couple blockchain traceability with a QR-triggered multilingual conversational LLM for non-technical consumers.

KEYWORDS

Blockchain, Agricultural Traceability, Smart Contract, Ethereum, Large Language Model, Conversational AI, QR Code, Food Safety, Multilingual Interface, Audit Trail, Consumer Transparency

I. INTRODUCTION

Food safety remains a critical global public health concern. According to the World Health Organization, approximately 600 million people suffer from foodborne illnesses annually, leading to around 420,000 deaths [1]. In India, which is the second-largest producer of fruits and vegetables globally, consumers are unable to confirm the source, use of pesticides, transport conditions, or quality certifications of the produce they buy. This lack of information creates situations where unsafe practices are not noticed at the consumer level.

Blockchain technology presents a promising solution to the problem of unclear supply chains. Its decentralized and append-only structure ensures that data cannot be changed, is secure, and provides clear records of the product's origin [2][3]. Ethereum further enhances this by using smart contracts that can automatically execute processes and enforce rules without the need for a trusted third party [4]. While several studies have shown that blockchain can work for tracking agricultural products on the blockchain [5][6][7], it still does not solve the issue of accessing this information for regular consumers—raw data on the blockchain is difficult for them to understand [8].

Recent research has started to tackle this issue by integrating large language models. Alves et al. [9] developed a chatbot powered by a retrieval-augmented generation model for blockchain-based traceability, while Neves et al. [10] suggested combining blockchain with retrieval-augmented generation and ReAct reasoning for better consumer interaction. However, these solutions do not support multiple languages or allow for easy access through QR codes—which are essential in the Indian agricultural retail environment.

This paper introduces FarmEcho, a system that combines Ethereum-based smart contracts with a conversational language model interface activated by a QR code, supporting three languages: English, Hindi, and Marathi. The main features of FarmEcho are: (i) smart contracts that are immutable and track changes over time; (ii) access to the language model via a QR code without needing to install an app; (iii) support for multiple languages in conversations; and (iv) detailed audit trails at the field level. To the best of our knowledge, FarmEcho is the first system that

combines blockchain traceability with a QR code-triggered multilingual conversational interface.

II. RELATED WORK

A. Blockchain-Based Agricultural Traceability

Researchers are studying Blockchain-based traceability because it provides a solution to the data integrity problems that exist in conventional supply chains. The decentralized ledger system introduced by Nakamoto [22] received further development from Buterin [4] who used Ethereum's smart contracts to create the technological foundation for agricultural traceability systems.

Wang et al. [2] developed a food supply chain traceability system on Ethereum through their implementation of smart contracts which produced immutable records that displayed verified product origins. Yang et al. [3] established a trustworthy blockchain system for fruits and vegetables by employing both on-chain and off-chain data storage methods to achieve data accuracy and system expansion. Salah et al. [5] created a multi-stakeholder model for agricultural supply chain registration, and Shahid et al. [6] presented a comprehensive agri-food solution covering certification, payment, and provenance tracking. The studies conducted in this research field use hybrid storage as their primary method by placing cryptographic hashes on the blockchain while keeping the actual data stored outside of the network. González-Amarillo et al. [13] studied IoT technology which acts as an extra monitoring instrument that works together with blockchain registration.

Kravenkit and So-In [15] proved that organizations can conduct product recalls more successfully through permanent records than they can through centralized systems which lack this capability. The systems which we examined all suffered from one main restriction because they could not provide consumers with access to data which remained unchanged. Tharatipyakul and Pongnumkul [7] found that consumer interfaces function as permanent elements which demand users to search through organized records which makes it hard for people who lack digital skills to use them. Tharatipyakul et al. [14] found that the Thai coffee traceability system stops users from participating because its interface design presents excessive challenges. Saidu et al. [8] and Misra et al. [21] found that agricultural traceability research needs more investigation into how consumers interact with accessible features and digital interfaces.

B. Large Language Models for Traceability Accessibility

Large Language Models (LLMs) create new methods to connect structured blockchain data with consumer requirements. Brown et al. [17] demonstrated that transformer-based models which include GPT-3 can perform complex reasoning tasks across multiple domains with minimal fine-tuning requirements. Touvron et al. [18] demonstrated that open-weight models like LLaMA can enable conversational AI without relying on proprietary APIs. Lewis et al. [19] introduced the RAG (Retrieval-Augmented Generation) method — fetching verified external knowledge when needed — which forms the basis of FarmEcho's approach to context injection.

Alves et al. [9] developed the most related work: an LLM chatbot for blockchain traceability using RAG, allowing users to ask natural language food safety questions without directly interacting with the blockchain. Neves et al. [10] expanded on this by combining blockchain, RAG, and ReAct reasoning to improve the accuracy of LLM responses on traceability data. Ferrão et al. [12] noted that consumer experience is still not a major focus in agricultural traceability research. Malik et al. [11] identified supply chain management as a key area for blockchain-anchored LLMs, but they also pointed out that there are few empirically tested implementations.

C. Research Gap

The existing literature follows three main paths: blockchain provides a foundation for immutable traceability [2][3][5][6]; LLMs offer natural-language accessibility [9][10][17][19]; and combining both is technically possible [9][10].

However, two important gaps remain. First, no system allows users to access an LLM through a QR code on physical packaging without needing to install an app — a key requirement in areas where digital literacy is a major barrier to adoption [7][14]. Second, no system supports multilingual options in Indian regional languages like Hindi or Marathi, even though most Indian consumers prefer these languages over English [21]. FarmEcho addresses

both of these issues within a single, tested framework.

III. SYSTEM ARCHITECTURE

A. Overview

FarmEcho is structured into three vertically integrated layers that work together seamlessly. The first layer is a blockchain-based data system that securely records supply chain information. The second layer includes servers that handle data retrieval and manage the interaction with large language models (LLMs). The third layer is a user-friendly interface that allows consumers to access food safety details by scanning a QR code with their smartphone camera. These layers combine raw data from the blockchain into easy-to-understand, conversational food safety information.

B. Supply Chain Stakeholders and Data Flow

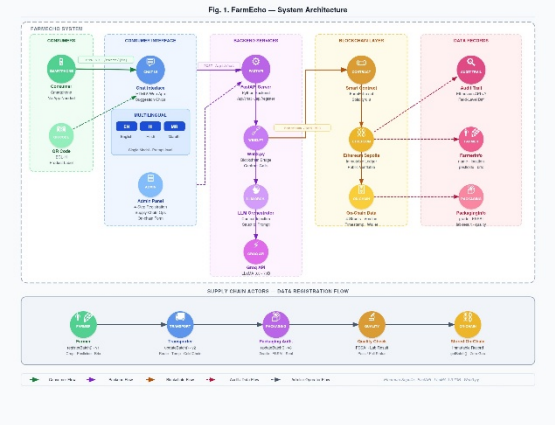
FarmEcho is designed to include four key participants in the supply chain. A farmer records details about their product, such as where it was grown, farming methods used, certifications held, planting and harvesting dates, use of pesticides, sugar content, and total yield. A transporter provides data about how the product is moved, including vehicle details, route, storage temperature, and whether cold chain conditions were followed — this information is stored as a simple yes or no and is locked in to prevent changes later without creating a record. A packaging authority shares information like the grade of the product, licensing details, lab test results, quality status, and best-before dates. Consumers only interact with the system through the conversational interface by scanning a QR code, and they never see or touch the blockchain directly.

C. Smart Contract Architecture

The FarmEcho smart contract in Solidity contains four data types which are structured as FarmerInfo and CropInfo and TransportInfo and PackagingInfo to store vital information. The system contains three primary functions. The registerBatch function stores the entire supply chain record and sends a timestamped event. The updateBatch function lets approved users change records, updates a version number, and keeps a complete history. The getBatch function allows users to retrieve information without making any changes. Encrypted digital signatures secure all updates which establish permanent records of each registration and update while enabling tracking of their original sources.

D. LLM Orchestration and Context Injection

The system uses Web3.py to obtain the complete batch record when a consumer scans a QR code. The system transforms this data into a natural language prompt which enables the LLM to comprehend the surrounding situation. The LLM API receives conversation history information from users. The model uses blockchain data as its contextual information instead of its built-in knowledge which helps produce correct answers. The prompt includes the consumer's preferred language which enables the model to generate accurate responses. The system implements the RAG method but enables direct data structure access without requiring an additional database.



IV. METHODOLOGY

A. Smart Contract Design

The FarmEcho smart contract operates on the Ethereum Sepolia network through its implementation in Solidity version 0.8. The system uses a main Batch struct which contains four smaller struct components named FarmerInfo and CropInfo and TransportInfo and PackagingInfo together with a registration date and a last-update date and a version number. The system uses a string-to-struct mapping method which enables O(1) retrieval by batch ID. The system provides access to four different functions. The registerBatch function checks for uniqueness, stores the batch data, and sends a BatchRegistered event. The updateBatch function changes existing record information while increasing version number and creating a BatchUpdated event. The getBatch function provides users with unmodifiable access to batch

data because it operates in a non-editable mode. The batchExists function checks system requirements to determine whether it should proceed with registration and update processes. The system Records all user write activities through caller private key which links each change to a specific Ethereum wallet address. The system stores Brix level data as an integer value which represents ten times the actual Brix level because Solidity does not enable usage of floating-point numbers.

B. Batch Registration Workflow

Registration process which requires an operator to complete A four-step web form that collects information about the farmer crop and transport method and packaging details. The frontend converts this information into a JSON payload and sends it to the FastAPI backend. The backend creates struct data with 30% additional gas costs which they sign to transmit the transaction while obtaining the transaction hash and block number and Etherscan link. Users can operate the system without needing to learn about blockchain technology or command-line interface operations.

C. QR Code Access Model

The system generates a QR code after batch registration which establishes a connection to a unique batch URL that Python's qrcode library produces with Error Correction Level H. The specified level enables up to 30% data loss while maintaining dependable scanning performance during actual operational usage. Users can access the consumer chat interface through their smartphone browser by scanning the QR code because it does not require them to install an app or create an account or enter their batch ID. Tharatipyakul and Pongnumkul discovered through their research [7] that the solution effectively addresses the primary challenge which they found.

D. LLM Context Injection Strategy

The backend system uses the getBatch function to obtain complete batch information in response to each consumer request. The system prompt which exists in structured form contains all struct fields as different parts which have their own specific labels. The language instruction together with conversation history increases the prompt which gets sent to the LLM API. The system uses RAG [19] technology to handle its data storage requirements because it processes prearranged material which has restricted access. The prompt requires the model to issue safety warnings when product quality assessments fall below acceptable levels and to determine product freshness based on harvest date and to explain Brix levels using everyday language and to create spoken dialogues instead of presenting raw data.

E. Multilingual Interface Design

The system enables multilingual support through system prompts which establish complete functionality without needing separate model instances or translation pipelines or fine-tuning processes. When a consumer selects a language (English, Hindi, or Marathi), that choice is added as a strict language instruction with each API call. The LLaMA 3.3 70B model, accessed via the Groq inference API, generates accurate responses in all three languages from a single model instance. The system enables users to obtain blockchain-verified Marathi answers about the pesticide spraying date by using the question "या फळावर किती दिवसांपूर्वी कीटकनाशक फवारले?" (How many days ago was pesticide sprayed?) without needing to install any application.

F. On-Chain Audit Trail and Field-Level Diff

The Etherscan API v2 delivers all contract associated transactions through its audit trail system. The contract's ABI enables decoding of each transaction input which reconstructs the exact struct field values submitted at that point in time. The method uses consecutive snapshots to identify field modifications which display which fields had their previous and current values changed. The audit event records six specific attributes which include action type, block number, UTC timestamp, wallet address, gas consumed, and complete field state at the time of the event. The system provides regulators with complete information about system modifications since it shows what changes occurred, when those changes happened, and who made them. The system records changes to all network activities for permanent storage because central systems permit selective log deletions.

V. IMPLEMENTATION

A. Deployment Environment

FarmEcho operates on the Ethereum Sepolia public testnet which replicates the mainnet testing

conditions to deliver precise gas cost assessments and permanent block finality. The system consists of approximately 1,200 Python code lines and 650 Solidity code lines and 2,400 HTML CSS and JavaScript code lines which are distributed among four frontend templates.

B. Smart Contract Deployment

The smart contract uses solcx for its compilation process while targeting Solidity version 0.8.0 and saving the ABI in a local directory. A signed transaction is required for contract deployment which needs to be sent through an Ethereum RPC provider. The contract address gets recorded in local storage and becomes accessible once the server initiates. Users can access all test transactions through the Sepolia Etherscan explorer because they exist in the public domain.

C. Test Batch Scenarios

Four test batches were registered which covered three different crop types. These tests included both safe and unsafe outcomes which Table III describes.

TABLE III — Test Batches Registered on FarmEcho Contract

Batch ID	Crop	Quality	Scenario Purpose
FTBATCH-2025-001	Alphonso Mango	✅ Passed	Safe — organic, zero pesticide
FTBATCH-2025-002	Sharad Seedless Grapes	❌ Failed	Unsafe — pesticide residue detected
FTBATCH-2025-003	Tomato (Hybrid F1)	✅ Passed	Safe — standard farming
FTBATCH-2025-004	Cherry Tomato	✅ Passed	Updated batch — audit trail testing

FTBATCH-2025-002 verified correct LLM safety warning behaviour; FTBATCH-2025-004 verified the field-level diff mechanism.

D. Gas Cost Analysis

Gas consumption across all contract operations is summarized in Table IV.

TABLE IV — Gas Consumption per Contract Operation

Operation	Gas Used	Approx. Cost (USD)*
Contract Deployment	3,241,500	~$0.19
registerBatch	932,908	~$0.05
updateBatch	274,707	~$0.02
getBatch (read)	0	Free

Estimated at 2 Gwei, ETH at $3,200. Mainnet costs vary with congestion.

The system performs its local view calls without requiring any gas costs during its read operations. The system updates existing storage slots through its updateBatch function which results in 70% gas savings because it does not create new storage slots.

E. Consumer and Admin Interface Design

The full web application enables users to access its user interface without needing to create an account or install any software. The LLM automatically generates an initial batch report with quality status and safety highlights when loaded. The system provides language buttons together with pre-written suggestion chips to help non-technical users complete their tasks. The admin panel provides a four-step form that automatically fills all fields with current on-chain information to complete the registration process.

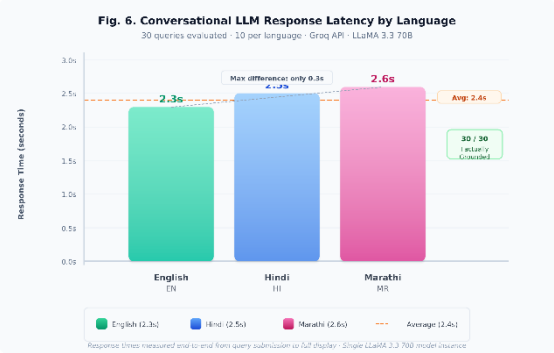
F. Multilingual Response Evaluation

Ten questions about safety evaluation, pesticide information, freshness, cold chain compliance, and quality certification were submitted in three languages for both a successful and an unsuccessful batch. The thirty answers all contained factual information without any fabricated field values. Safety warnings about the unsuccessful batch were correctly triggered in all three languages. The average response times were 2.3 seconds for English, 2.5 seconds for Hindi, and 2.6 seconds for Marathi.

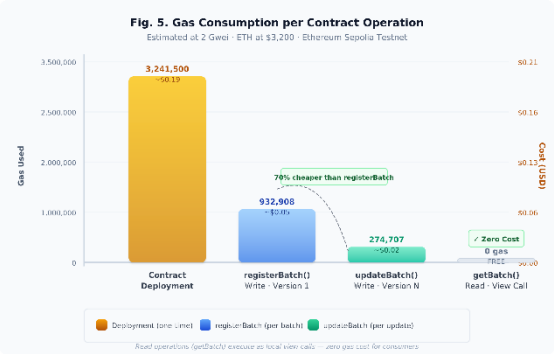
VI. RESULTS AND DISCUSSION

A. System Performance Summary

The assessment of FarmEcho measured its gas cost efficiency and its ability to provide quality conversational answers while assessing its usability for customers. The registration costs amounted to $0.05 per batch and $0.02 per update which smallholder farmers find to be affordable. The system provides free access to reading operations which allows customers to obtain information without paying any transaction fees. The system delivered a response time of 2.4 seconds in three languages which matched standard web application performance.



The thirty multilingual evaluation queries produced accurate results because all responses contained factual information without any incorrect values. The LLM issued safety alerts about the unsuccessful batch in all three languages. The results demonstrate that the context injection method restricts system outputs to authenticated blockchain information instead of relying on pre-existing knowledge.



B. Comparison with Existing Commercial Systems

The commercial platforms IBM Food Trust, Transparent Path, Ripe.io, TraceX, Agri-Ledger, GrainChain, and HARA present traceability information through their websites and applications which need users to have navigation skills and technical expertise. The platforms enable blockchain-based traceability but they do not include a conversational AI system that understands natural language queries from users who speak Indian regional languages. The two systems IBM Food Trust and TraceX permit users to scan QR codes but direct them to unproductive data displays instead of engaging experiences. The two platforms GrainChain and AgriChain enable transactions between operators but they do not offer direct access for consumers. The platform HARA studies smallholder access in Indonesia but it does not support users with its conversational or multilingual features. FarmEcho distinguishes itself from other platforms by providing users with a framework that enables them to access QRs without installations while using multilingual conversational AI and field-based auditing capabilities.

C. Comparison with Research Systems

Alves et al. [9] created an LLM chatbot which tracks blockchain data through RAG but their system lacks both QR activation and multiple language support and audit tracking functions. The framework proposed by Neves et al. [10] used ReAct reasoning but their system lacked both multilingual support and zero-installation access for users. FarmEcho provides both solutions through its single deployable system.

D. Discussion

The primary objective of FarmEcho gets accomplished through its ability to provide non-technical users with access to blockchain traceability data without requiring them to install an application or possess technical skills or understand the English language. The field-level audit trail provides a solution for a problem which both commercial and academic systems have failed to resolve. The system shows regulators the complete history of changes which includes details about the date and time of each modification and the identity of the user who executed the encrypted modification. Food safety enforcement requires this particular information.

The current testnet deployment creates a limitation because it fails to deliver the economic security protections which the mainnet provides. The upcoming research will concentrate on three areas which include mainnet deployment and IoT sensor integration for automatic data collection and large-scale tests with Indian agricultural communities.

VII. CONCLUSION

This paper presented FarmEcho, which serves as an agricultural traceability system that operates on blockchain technology and uses Ethereum smart contracts with a QR-based LLM interface. The system enables secure supply chain record verification while permitting consumers to retrieve product details through natural language queries that do not require application installation. The system offers multilingual support which includes English, Hindi, and Marathi to enhance accessibility options.

The assessment demonstrated that the system could function effectively because it required only minimal costs to register users and it charged no fees for reading content while it produced correct answers to multiple language requests. FarmEcho establishes a new standard through its unique system which combines blockchain traceability with QR-based consumer access and conversational AI capabilities into a single framework.

The upcoming tasks will require deploying the mainnet system which needs to integrate IoT sensors for automatic data collection while extending its capabilities to support multiple regional languages and connect with FSSAI's national food safety system.

REFERENCES

[1] World Health Organization, "Food Safety," WHO Fact Sheet, Apr. 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/food-safety

[2] L. Wang et al., "Smart Contract-Based Agricultural Food Supply Chain Traceability," *IEEE Access*, vol. 9, pp. 9296–9307, 2021, doi: 10.1109/ACCESS.2021.3050112.

[3] X. Yang, M. Li, H. Yu, M. Wang, D. Xu, and C. Sun, "A Trusted Blockchain-Based Traceability System for Fruit and Vegetable Agricultural Products," *IEEE Access*, vol. 9, pp. 36282–36293, 2021, doi: 10.1109/ACCESS.2021.3062845.

[4] V. Buterin, "A Next-Generation Smart Contract and Decentralized Application Platform," Ethereum White Paper, 2014. [Online]. Available: https://ethereum.org/en/whitepaper

[5] K. Salah, M. H. U. Rehman, N. Nizamuddin, and A. Al-Fuqaha, "Blockchain-Based Soybean Traceability in Agricultural Supply Chain," *IEEE Access*, vol. 7, pp. 73295–73305, 2019, doi: 10.1109/ACCESS.2019.2918000.

[6] A. Shahid et al., "Blockchain-Based Agri-Food Supply Chain: A Complete Solution," *IEEE Access*, vol. 8, pp. 69230–69243, 2020, doi: 10.1109/ACCESS.2020.2986257.

[7] A. Tharatipyakul and S. Pongnumkul, "User Interface of Blockchain-Based Agri-Food Traceability Applications: A Review," *IEEE Access*, vol. 9, pp. 82909–82929, 2021, doi: 10.1109/ACCESS.2021.3085982.

[8] Y. Saidu et al., "Convergence of Blockchain, IoT, and AI for Enhanced Traceability Systems: A Comprehensive Review," *IEEE Access*, vol. 13, pp. 16838–16865, 2025, doi: 10.1109/ACCESS.2025.3528035.

[9] J. Alves et al., "LLM Based Chatbot for Farm-to-Fork Blockchain Traceability Platform," *Applied Sciences*, vol. 14, no. 19, p. 8856, Oct. 2024, doi: 10.3390/app14198856.

[10] R. Neves et al., "Using LLMs to Bridge the Gap Between Consumers and Blockchain on an Agro-Food Traceability Platform," *Procedia Computer Science*, Mar. 2025, doi: 10.1016/j.procs.2025.04.831.

[11] S. Malik et al., "Blockchain for LLMs — Applications, Challenges, and Framework," *Expert Systems with Applications*, Jan. 2026, doi: 10.1016/j.eswa.2026.00014.

[12] R. Ferrão et al., "GenAI and Blockchain to Coordinate Agricultural Supply Chains," *Frontiers in Sustainable Food Systems*, vol. 9, Jul. 2025, doi: 10.3389/fsufs.2025.1591350.

[13] C. A. González-Amarillo et al., "An IoT-Based Traceability System for Greenhouse Seedling Crops," *IEEE Access*, vol. 6, pp. 67528–67535, 2018, doi: 10.1109/ACCESS.2018.2877293.

[14] A. Tharatipyakul, S. Pongnumkul et al., "Blockchain-Based Traceability System From the Users' Perspective: A Case Study of Thai Coffee Supply Chain," *IEEE Access*, vol. 10, pp. 98783–98802, 2022, doi: 10.1109/ACCESS.2022.3206860.

[15] S. Kravenkit and C. So-In, "Blockchain-Based Traceability System for Product Recall," *IEEE Access*, vol. 10, pp. 95132–95150, 2022, doi: 10.1109/ACCESS.2022.3204750.

[17] T. B. Brown et al., "Language Models are Few-Shot Learners," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 1877–1901, 2020.

[18] H. Touvron et al., "LLaMA: Open and Efficient Foundation Language Models," *arXiv preprint*, arXiv:2302.13971, Feb. 2023.

[19] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Proc. NeurIPS*, vol. 33, pp. 9459–9474, 2020.

[21] N. N. Misra et al., "IoT, Big Data, and Artificial Intelligence in Agriculture and Food Industry," *IEEE Internet of Things Journal*, vol. 9, no. 9, pp. 6305–6324, May 2022, doi: 10.1109/JIOT.2020.2998584.

[22] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," White Paper, Oct. 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf

FitAI: Интеллектуальная система анализа техники упражнений на основе компьютерного зрения и машинного обучения

Токонов Айдар
Кыргызско-Турецкий Университет Манас
Отдел компьютерной инженерии
Бишкек, Кыргызская Республика
aydarmaratov17@gmail.com

Жумабаева Чинара
Кыргызско-Турецкий Университет Манас
Отдел компьютерной инженерии
Бишкек, Кыргызская Республика
chinara.jumabaeva@manas.edu.kg

***Аннотация*—В статье представлена система FitAI — интеллектуальный инструмент для автоматического анализа техники выполнения физических упражнений в режиме реального времени на основе компьютерного зрения и глубокого обучения. Система использует библиотеку MediaPipe для извлечения 33 ключевых точек тела и вычисляет 78-мерный вектор признаков (66 координат + 12 суставных углов). Для классификации упражнений применяется двунаправленная рекуррентная сеть с долгой краткосрочной памятью (BiLSTM, от англ. Bidirectional Long Short-Term Memory), анализирующая последовательности из 30 кадров. Оценка техники выполняется с помощью правилобазированного алгоритма на основе суставных углов. Точность классификации на тестовой выборке составила 99,73%. Система поддерживает четыре класса упражнений: приседания, отжимания, сгибание рук на бицепс и жим над головой.**

Ключевые слова—компьютерное зрение; машинное обучение; BiLSTM; MediaPipe; анализ техники упражнений; распознавание активности человека; фитнес

I. ВВЕДЕНИЕ

Физическая активность является ключевым фактором сохранения здоровья. Тем не менее статистика свидетельствует о серьёзной проблеме: 72% начинающих посетителей тренажёрного зала выполняют базовые упражнения с нарушением техники, не осознавая этого. Каждая третья спортивная травма в зале вызвана не перегрузкой, а неправильной биомеханикой движения [1].

Единственным надёжным источником коррекции техники традиционно остаётся персональный тренер, стоимость услуг которого составляет 20–100 долларов за сессию. Для большинства людей — особенно в странах с развивающейся экономикой — это попросту недоступно. Таким образом, складывается парадокс: фитнес-контент стал массовым и бесплатным, а качественная обратная связь по технике по-прежнему остаётся привилегией.

Современные достижения в области компьютерного зрения и рекуррентных нейронных сетей открывают возможность автоматизировать эту функцию тренера. Рынок интеллектуальных фитнес-приложений демонстрирует устойчивый рост, однако большинство из них не предоставляет обратной связи по биомеханике движений [26]. Настоящая статья описывает систему FitAI, которая в режиме реального времени распознаёт тип упражнения и оценивает корректность его выполнения, предоставляя конкретную текстовую обратную связь без участия человека.

II. ОБЗОР СВЯЗАННЫХ РАБОТ

Задача оценки техники упражнений лежит на пересечении двух направлений: оценка позы тела человека (HPE, от англ. Human Pose Estimation) и распознавание физической активности (HAR, от англ. Human Activity Recognition).

A. A. Оценка позы тела человека

Работа Cao et al. (2017) [2] заложила основы многоперсонного определения ключевых точек тела в реальном времени на основе Part Affinity Fields. Ранее Wei et al. [10] предложили архитектуру Convolutional Pose Machines, продемонстрировав эффективность последовательной свёрточной обработки. Google MediaPipe [3] предложил облегчённую мобильную реализацию без GPU [13], способную определять 33 ключевые точки тела на смартфоне.

B. B. Распознавание физической активности

Архитектуры долгой краткосрочной памяти (LSTM) [4] и BiLSTM [5] зарекомендовали себя как эффективный инструмент для анализа временных последовательностей кинематических данных [11]. Обзор Aggarwal и Ryoo [12] показал превосходство подходов на основе скелетных данных. Ключевой методологической основой FitAI послужила работа Riccio (2024) [6], продемонстрировавшая сочетание MediaPipe + BiLSTM для классификации фитнес-упражнений.

C. C. Ограничения существующих решений

Коммерческие системы (Google Fit, Apple Fitness+) фиксируют физическую активность, однако не оценивают биомеханическую корректность движений. CNN + голосование по кадрам [19] игнорирует временной контекст; LSTM на сырых координатах чувствителен к позиции камеры; однокадровая классификация теряет динамику движения. Предложенный BiLSTM + суставные углы устраняет все три ограничения одновременно.

III. АРХИТЕКТУРА СИСТЕМЫ

FitAI реализован как многокомпонентная платформа: серверная часть на ASP.NET Core, мобильное приложение на Flutter, AI-сервис на FastAPI, хранилище данных PostgreSQL. Все компоненты контейнеризированы с использованием Docker [13]. Конвейер обработки данных включает шесть этапов:

1. Видеокадр с камеры передаётся в модуль извлечения признаков.
2. MediaPipe Pose выделяет 33 ключевые точки тела; используется 22 из них (индексы 11–32, лицо исключено).
3. Формируется 78-мерный вектор: 66 координат (x, y, visibility) + 12 суставных углов.
4. BiLSTM классифицирует тип упражнения по последовательности из 30 кадров (~1 сек).
5. Правилобазированный модуль рассчитывает оценку техники по суставным углам.
6. Пользователь получает конкретное текстовое сообщение или подтверждение корректности.

IV. ИЗВЛЕЧЕНИЕ ПРИЗНАКОВ

Центральным решением системы является использование суставных углов в качестве основного признака, а не сырых координат [14]. Каждая из 22 ключевых точек предоставляет значения x, y, visibility, итого 66 координатных признаков. Дополнительно вычисляется 12 суставных углов (локти, плечи, бёдра, колени, лодыжки, запястья) по теореме косинусов:

$$\theta = \arccos\left((b^2 + c^2 - a^2) / 2bc \right) \quad (1)$$

где a, b, c — длины сторон треугольника, образованного тремя смежными ключевыми точками [9]. Суставные углы обеспечивают инвариантность к положению камеры и расстоянию до объекта. Итоговый вектор: 66 + 12 = 78 признаков на кадр.

V. МОДЕЛЬ КЛАССИФИКАЦИИ BILSTM

Упражнение является движением во времени, а не статической позой. Для захвата временного контекста применяется BiLSTM, анализирующая последовательность из 30 кадров (≈1 секунда при 30 FPS [22]) как в прямом, так и в обратном направлении. Для сглаживания предсказаний используется скользящее окно с 50% перекрытием [15].

Архитектура модели (вход 30 × 78):

— BiLSTM (256 единиц) → Dropout (0.3) — локальные паттерны движения;

— BiLSTM (128 единиц) → Dropout (0.3) — глобальные временные зависимости;

— Dense (64, ReLU) → Dropout (0.3);

— Dense (4, Softmax) — выход: Squat / Push-up / Bicep curl / Shoulder press.

Оптимизатор Adam (lr=0.001) [20], размер батча 32, ранняя остановка при patience=8. Слои Dropout применяются для регуляризации [21].

VI. ОЦЕНКА ТЕХНИКИ ВЫПОЛНЕНИЯ

После классификации активируется правилобазированный оценщик. Гибридный подход (нейросеть + биомеханические правила) доказал эффективность в анализе движений спортсменов [16]. Алгоритм: (1) аккумулируются последние 10 кадров; (2) усредняются 12 суставных углов; (3) проверяются биомеханические правила; (4) score = 1 − (нарушенный вес / общий вес); (5) при score ≥ 0.65 техника корректна.

Пример: для приседания — угол колена ≤ 100° (вес 0.40), угол бедра ≤ 120° (вес 0.35); отжимание — угол локтя ≤ 100° (вес 0.40); жим над головой — угол локтя ≥ 155° (вес 0.45). Пример обратной связи: «Согните колени глубже — угол не должен превышать 90–100°».

VII. ОБУЧАЮЩИЕ ДАННЫЕ

Датасет сформирован из трёх источников: синтетические видео InfiniteRep v1.0 [17], реальные видео из Kaggle Gym Exercise Dataset [18] и веб-краулинг. Разбивка: обучение 70% / валидация 15% / тест 15%. Состав датасета представлен в Табл. I.

ТАБЛ. I. СОСТАВ ОБУЧАЮЩЕГО ДАТАСЕТА

Класс упражнения	Кадров	Доля, %
Приседание (Squat)	53 149	35.3
Отжимание (Push-up)	38 293	25.4
Сгибание рук (Bicep curl)	32 099	21.3
Жим над головой (Shoulder press)	27 119	18.0
Итого	**~150 660**	**100**

VIII. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Модель достигла точности 99.73% на тестовой выборке и 99.87% на валидационной. Из 1506 тестовых образцов допущено лишь 4 ошибки: 2 — в классе Bicep curl (спутаны с Squat), 1 — Push-up→Bicep curl, 1 — Shoulder press→Squat. Метрики по классам представлены в Табл. II.

ТАБЛ. II. МЕТРИКИ КЛАССИФИКАЦИИ (ТЕСТОВАЯ ВЫБОРКА)

Класс	Precision	Recall	F1	N
Приседание	0.99	1.00	1.00	531
Отжимание	1.00	1.00	1.00	383
Сгибание рук	1.00	0.99	1.00	321
Жим над головой	1.00	1.00	1.00	271
Macro avg.	**1.00**	**1.00**	**1.00**	**1506**

Суставные углы устраняют чувствительность к позиции камеры; окно в 30 кадров даёт достаточный временной контекст; двунаправленная обработка учитывает полный цикл движения. Среднее время обработки кадра — менее 33 мс, что обеспечивает работу при 30 FPS [22].

IX. ОБСУЖДЕНИЕ И ОГРАНИЧЕНИЯ

Предложенный метод BiLSTM + суставные углы единственный среди рассмотренных обеспечивает одновременно инвариантность к положению камеры и сохранение временного контекста. Текущие ограничения: библиотека ограничена четырьмя классами упражнений; точность MediaPipe снижается при перекрытии тела или экстремальных ракурсах [23]; пороги правил не учитывают антропометрию пользователя. Направления развития: расширение до 15+ классов, персонализация порогов [24], применение 3D-оценки позы [25].

X. ЗАКЛЮЧЕНИЕ

В статье представлена система FitAI — автоматическое распознавание и оценка техники физических упражнений в режиме реального времени. Ключевые результаты:

— полный конвейер от видеокадра до интерпретируемой обратной связи;

— 78-мерное инвариантное представление признаков (координаты + углы);

— точность классификации 99.73% на ~150 660 кадрах четырёх классов;

— гибридная архитектура (нейросеть + правила) обеспечивает точность и понятную обратную связь.

FitAI демонстрирует, что персональный тренер по технике может быть воспроизведён автоматически при нулевых дополнительных расходах для пользователя.

Список литературы

[1] [1] U. Aasa, I. Svartholm, F. Andersson, and L. Berglund, "Injuries among weightlifters and powerlifters: a systematic review," Br. J. Sports Med., vol. 51, pp. 211–219, 2017.

[2] [2] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in Proc. IEEE CVPR, 2017, pp. 7291–7299.

[3] [3] C. Lugaresi et al., "MediaPipe: A framework for building perception pipelines," arXiv:1906.08172, 2019.

[4] [4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[5] [5] M. Schuster and K. Paliwal, "Bidirectional recurrent neural networks," IEEE Trans. Signal Process., vol. 45, no. 11, pp. 2673–2681, 1997.

[6] [6] R. Riccio, "Real-time fitness exercise classification and counting from video frames," arXiv:2411.11548, Univ. of Pavia, 2024.

[7] [7] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.

[8] [8] R. Szeliski, Computer Vision: Algorithms and Applications, 2nd ed. New York: Springer, 2022.

[9] [9] G. Strang, Linear Algebra and Its Applications, 4th ed. Cengage Learning, 2006.

[10] [10] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in Proc. IEEE CVPR, 2016, pp. 4724–4732.

[11] [11] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM," Neural Netw., vol. 18, no. 5–6, pp. 602–610, 2005.

[12] [12] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: a review," ACM Comput. Surv., vol. 43, no. 3, pp. 1–43, 2011.

[13] [13] D. Merkel, "Docker: lightweight Linux containers," Linux J., vol. 239, p. 2, 2014.

[14] [14] D. C. Luvizon, H. Tabia, and D. Picard, "Human pose regression by combining indirect part detection," Comput. Graph., vol. 85, pp. 15–22, 2019.

[15] [15] J. Yang et al., "Deep CNN on multichannel time series for HAR," in Proc. IJCAI, 2015, pp. 3995–4001.

[16] [16] H. Pirsiavash and D. Ramanan, "Assessing the quality of actions," in Proc. ECCV, 2014, pp. 556–571.

[17] [17] C. Doersch and A. Zisserman, "Sim2Real transfer learning for 3D pose estimation," arXiv:1907.02499, 2019.

[18] [18] Kaggle Gym Exercise Dataset. [Online]. Available: https://www.kaggle.com/datasets/hasyimabdillah/workoutfitness-video

[19] [19] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition," in Proc. NeurIPS, 2014, pp. 568–576.

[20] [20] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in Proc. ICLR, 2015, arXiv:1412.6980.

[21] [21] N. Srivastava et al., "Dropout: a simple way to prevent neural networks from overfitting," J. Mach. Learn. Res., vol. 15, pp. 1929–1958, 2014.

[22] [22] J. Shotton et al., "Real-time human pose recognition in parts from single depth images," in Proc. IEEE CVPR, 2011, pp. 1297–1304.

[23] [23] Y. Cheng et al., "Occlusion-aware networks for 3D human pose estimation in video," in Proc. IEEE ICCV, 2019, pp. 723–732.

[24] [24] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "RMPE: regional multi-person pose estimation," in Proc. IEEE ICCV, 2017, pp. 2334–2343.

[25] [25] D. Pavllo et al., "3D human pose estimation in video with temporal convolutions," in Proc. IEEE CVPR, 2019, pp. 7753–7762.

[26] [26] Grand View Research, "Fitness App Market Size, Share & Trends Analysis Report," 2023. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/fitness-app-market

Developing an AI-Based Assistant for Selecting Educational Direction and Exam Preparation According to Student Ability

Zhoomart Suranbaev
Department of Computer Engineering
Kyrgyz-Turkish Manas University
Bishkek, Kyrgyz Republic
2204.01014@manas.edu.kg

Chinara Jumabaeva
Department of Computer Engineering
Kyrgyz-Turkish Manas University
Bishkek, Kyrgyz Republic
chinara.jumabaeva@manas.edu.kg

***Abstract*—Personalized learning is becoming increasingly important because students differ in prior knowledge, learning speed, and academic ability. A single preparation strategy is often inefficient when students must select an educational direction and prepare for entrance or placement exams. This paper presents a scalable AI-based assistant that analyzes test attempts, identifies weak topics, recommends study actions, ranks educational programs, and answers user questions through a RAG-style grounded assistant. The prototype is implemented with a React and TypeScript frontend, a FastAPI backend, and a SQLModel data layer; SQLite is used for local development and PostgreSQL is supported for deployment. The machine-learning layer includes a Random Forest topic diagnosis model, BKT-aware adaptive blending, and a PyTorch/LambdaRank model for program or institution ranking. The RAG component builds a runtime corpus from institution, program, and test records, retrieves top ranked sources with profile-aware lexical scoring, and then returns citation-backed answers through an LLM when available or a deterministic local fallback when the external model is unavailable. Internal validation shows that the Random Forest topic model reaches an AUC of 0.8803, the adaptive topic blend reaches an AUC of 0.9127, and the ranking model reaches an NDCG at 10 of 0.8268. The results suggest that a hybrid architecture can support personalized exam preparation and educational direction selection in low-data settings while remaining interpretable and portable across regions and exam systems.**

***Index Terms*—artificial intelligence, personalized learning, educational recommendation, Random Forest, Bayesian knowledge tracing, LambdaRank, retrieval-augmented generation, grounded assistant**

I. Introduction

Choosing an educational direction and preparing for entrance or placement exams are high-stakes tasks for students. Applicants need to compare programs, estimate whether their current ability is sufficient, and choose a preparation strategy that matches their weak topics rather than following a generic test-bank path [1].

Traditional preparation usually follows a one-size-fits-all approach. Students solve many questions but receive limited explanation about which topics are genuinely weak. Program selection is also often reduced to manual comparison of admission requirements. This creates a gap between student ability, preparation strategy, and realistic educational direction.

The proposed assistant addresses this gap through a hybrid architecture. Machine-learning models estimate weak topics and ranked program fit, while domain rules preserve hard constraints such as score thresholds, location preference, and target field. The assistant therefore acts not only as a predictor but also as an explainable decision-support tool.

A. Contributions

The paper contributes: (1) a deployed AI assistant architecture for exam preparation and educational direction selection; (2) a topic diagnosis model for sparse student histories; (3) an implementation-faithful RAG-style grounded assistant over local institution, program, and test data; (4) a LambdaRank-based ranking model; and (5) a hybrid scoring layer that combines learned predictions with expert rules.

II. Related Work

A. Knowledge tracing and personalization

Bayesian Knowledge Tracing models mastery as a latent state updated after each observation [3]. Deep Knowledge Tracing later demonstrated the use of recurrent neural networks for student sequence modeling [2]. Recent work extends knowledge tracing with graph-based representations and other structured approaches [5]. Hybrid approaches also combine latent-factor models with knowledge tracing to predict individual differences in learning [8]. These approaches are powerful, but many require long interaction histories that are not always available in regional educational platforms.

B. Recommendation under cold start

Educational recommenders frequently face cold-start and sparsity problems. Context-aware, hybrid, and learning-network recommenders remain useful because they combine student attributes, content information, and domain rules when user-item interaction data is limited [4], [6], [7], [18], [19]. Neural collaborative filtering and pairwise ranking methods show the value of learned ranking, but they usually benefit

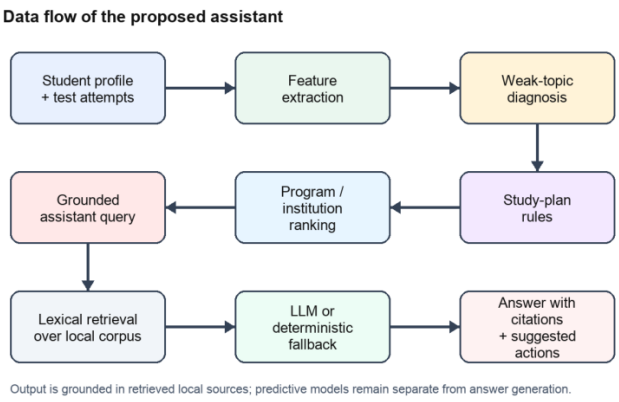


Fig. 1. Data flow of the proposed assistant.

from larger interaction logs than a new educational prototype can provide [20], [21]. For institution selection, ranking metrics such as NDCG are more appropriate than binary accuracy because the application displays ordered recommendation lists [10], [11].

C. Grounded generation and RAG

Retrieval-Augmented Generation connects a language model with an external memory so that generated answers are grounded in retrieved documents rather than only in model parameters [9]. REALM and DPR further show the importance of retrieval for knowledge-intensive and open-domain question answering [16], [22]. In the proposed assistant, this idea is implemented as a lightweight domain-specific pipeline: the retriever is lexical and rule-weighted rather than dense, but the system still follows the retrieve-then-generate pattern and can later be upgraded to BM25 or dense retrieval [17].

III. System Architecture

The system uses a three-layer architecture. The client is implemented in React and TypeScript. The backend uses FastAPI and exposes endpoints for tests, study planning, weak-topic prediction, educational recommendation, and the `/support/assistant` grounded assistant. The data layer stores users, test attempts, questions, institutions, programs, and tutors through SQLModel; the local prototype uses SQLite and the deployment configuration supports PostgreSQL. The current prototype catalog contains 74 institution records, 1493 program records, and 13 seeded tests.

The workflow is organized as a controlled pipeline: (1) collect profile and attempt history; (2) compute topic and behavioral features; (3) predict topic weakness; (4) build a rule-based topic study plan; (5) rank institutions; (6) blend model outputs with rules; and (7) answer user questions through a RAG-style grounded assistant that returns citations.

After each test attempt, the backend updates the student's profile and computes topic-level probabilities. For educational recommendation, it evaluates program constraints, obtains a learned ranking score, and blends that score with a rule-based explanation. The frontend can then show both the recommendation and the reason behind it.

IV. RAG-Style Grounded Assistant

A. Corpus construction

At request time, the `/support/assistant` endpoint builds a local corpus from three database sources. Each institution becomes a document containing city, major, institution type, address, phone, minimum score, tuition, exams, source name, and a short list of programs. Each ETL program becomes a separate document with degree, faculty, language, duration, admission requirements, exams, and source metadata. Each seeded test becomes a document with subject, topic list, and question count. This produces 1580 runtime documents in the current local database.

B. Lexical retrieval and profile-aware scoring

The retriever normalizes punctuation, removes common multilingual stopwords, and expands domain terms for IT, journalism/media, and military education, including transliterated forms. The latest user question is treated as the main intent, while profile context such as target major, preferred city, and average score is only a weak tie-breaker. This avoids the case where an old profile preference hides a new question such as journalism or military education.

$$QScore(q,d) = 4|Q \cap D| + 6|Q \cap T| + 3\,\mathrm{Substr}(Q,d) + 5\,\mathrm{Substr}(Q,T). \tag{1}$$

$$RScore(q,p,d) = QScore(q,d) + 0.5\min(|P \cap D|, 3) + B_{subset} + B_{title}. \tag{2}$$

$$TopK(q) = \arg\mathrm{top}_k\, RScore(q,p,d), \quad k = 3. \tag{3}$$

C. Generation and fallback

If `DEEPSEEK_API_KEY` is configured, the selected documents and profile context are sent to DeepSeek Chat Completions with a system instruction to answer only from the retrieved context and return a JSON object containing answer, citations, and suggested actions. If the external call is unavailable or no API key is configured, the endpoint returns a deterministic local answer from the same retrieved documents. Therefore the deployed module is accurately described as a RAG-style grounded assistant, not as a full trainable neural RAG model.

V. Topic Diagnosis Model

A. Feature representation

The topic model estimates $P(correct \mid student, topic)$. Its features include average score, tests taken, last score, subject, topic, question difficulty, topic attempt count, mastery estimate, and forgetting factor. Categorical fields are one-hot encoded, and the model is trained as a Random Forest

classifier. The enhanced feature set allows the assistant to distinguish between a difficult topic, a random mistake, and repeated evidence of weakness.

$$x_{s,t} = [avg_score, tests_taken, last_score, subject, topic, difficulty, attempt_count, mastery, forgetting]. \quad (4)$$

$$P_{correct}(s,t) = RF(x_{s,t}). \quad (5)$$

$$WeakTopics(s) = TopK_t(1 - P_{blend}(s,t)). \quad (6)$$

B. Adaptive blending

At inference time, the assistant combines population-level predictions with student-specific evidence. With no observations, the Random Forest estimate is used. With limited observations, observed accuracy is blended in cautiously. With more observations, the system shifts toward mastery-based evidence using a sigmoid weighting function. This makes the model useful during cold start and more personalized after practice accumulates.

$$w(n) = \frac{1}{1 + \exp(-0.5(n-5))}. \quad (7)$$

$$P_{blend}(s,t) = (1 - w(n))P_{RF}(s,t) + w(n)P_{mastery}(s,t), \quad n \geq 3. \quad (8)$$

VI. University Ranking Model

A. Problem formulation

Institution selection is modeled as a ranking task. The PyTorch model [12] receives 11 features: average score, city match, major match, minimum score, tuition, exam count, score gap, tests taken, score trend, topic accuracy, and program count. The network has layers 11–32–16–8–1 with batch normalization [13], ReLU activations, and dropout [14].

B. Hybrid scoring

The learned institution score is blended with a rule-based score that evaluates readiness, major alignment, city match, tuition feasibility, and exam complexity. In the deployed endpoint, rules receive the larger weight, while the ML model adjusts the ordering among close candidates. This prevents recommendations that violate basic admission logic while still using behavioral signals such as score trend.

$$score_gap(s,u) = avg_score_s - min_score_u. \quad (9)$$

$$RuleScore = 0.34R + 0.26M + 0.16C + 0.12T + 0.12E. \quad (10)$$

$$FinalScore(s,u) = 0.25MLRank(s,u) + 0.75RuleScore(s,u). \quad (11)$$

TABLE I
Data Sources and Current System Coverage

Component	Value
Institution records	74
Program records	1493
Seeded practice tests	13
Stored test attempts	1173
Unique users	31
Runtime RAG documents	1580

TABLE II
Model Summary and Validation Metrics

Model	Metric
Random Forest topic model	AUC = 0.8803
Adaptive RF + mastery blend	AUC = 0.9127
LambdaRank ranking model	NDCG@10 = 0.8268
Final ranking output	25% ML + 75% rules

VII. Experimental Results

The evaluation follows two complementary protocols. Topic diagnosis is evaluated as a probabilistic prediction task using AUC, accuracy, precision, recall, and F1-score. Institution recommendation is evaluated as a ranking task using NDCG@5, NDCG@10, and NDCG@20, because the user receives an ordered shortlist rather than a binary decision.

In the deployed workflow, each model has a separate responsibility. The Random Forest model estimates topic-level correctness probabilities and is used to prioritize weak topics. The BKT-aware blend adjusts these estimates as repeated attempts accumulate. The PyTorch/LambdaRank model orders institutions or programs for the shortlist, while the rule layer keeps admission constraints interpretable. The RAG-style assistant is not used to predict scores; it answers explanatory questions from retrieved local sources and returns citations.

$$DCG@k = \sum_{i=1}^{k} \frac{2^{rel_i} - 1}{\log_2(i+1)}. \quad (12)$$

$$NDCG@k = \frac{DCG@k}{IDCG@k}. \quad (13)$$

The Random Forest topic model reached 0.8803 AUC in offline validation. The adaptive topic blend improved the result to 0.9127 AUC by incorporating observation-based mastery. This supports the assumption that population-level patterns are useful at cold start, while direct student evidence becomes more valuable with repeated attempts.

For educational recommendation, the LambdaRank model reached NDCG@5 of 0.8512, NDCG@10 of 0.8268, and NDCG@20 of 0.7934. Because applicants usually inspect only a short list, NDCG@5 and NDCG@10 are more relevant than global classification metrics. The final hybrid score is designed to keep these rankings stable and interpretable.

VIII. Deployment

The backend loads model artifacts lazily and caches them for inference. The FastAPI service layer [15] loads the topic model serialized with joblib and the institution model stored

as a PyTorch state dictionary. The `/support/assistant` endpoint builds its retrieval corpus at request time from Institution, InstitutionProgram, and Test records, then returns answer text, citations, and suggested actions.

Cold-start handling is explicit. New users receive RF-based topic estimates and rule-dominant educational recommendations. As attempts accumulate, the system incorporates observed accuracy, score trend, and topic-level mastery.

IX. Discussion and Limitations

The results are consistent with knowledge-tracing literature. Classical Bayesian Knowledge Tracing models student mastery as a latent state updated after observations [3], while Deep Knowledge Tracing and graph-based variants show that sequential and structured histories can improve prediction [2], [5]. In this prototype, the strongest practical result is that a simpler Random Forest model combined with BKT-aware mastery evidence reaches high AUC even when student histories are sparse. This supports the use of lighter, interpretable models in early-stage educational systems.

The recommendation results also agree with prior work on educational and hybrid recommender systems. Surveys and applied studies show that context, content, and domain constraints are valuable when interaction data is limited [4], [6], [18], [19]. In this paper, the rule layer preserved explicit constraints such as score threshold, field match, and location preference, while LambdaRank adjusted the ordering among close candidates. This adds evidence that hybrid ranking can be more useful than a pure neural model in low-data educational recommendation settings.

The ranking evaluation follows information-retrieval practice. NDCG rewards correct ordering near the top of a list and is therefore aligned with the user experience of inspecting only a short shortlist [10], [11]. The obtained NDCG@10 result indicates that the ranking model learned a useful ordering signal, but it should be interpreted as internal validation because relevance labels are partly rule-derived.

The RAG-style assistant follows the same direction as retrieval-augmented generation and open-domain question answering work [9], [16], [22]. The practical result in this paper is narrower but still useful: retrieved local sources allow the assistant to return citation-backed answers and a deterministic fallback when an external LLM is unavailable. Unlike full neural RAG, the current retriever is lexical and rule-weighted; this differs from dense retrieval results and should be improved with BM25 or embeddings in future versions [17].

The main limitations are training data, retrieval evaluation, and external validity. Current model evaluation relies on seeded interactions and rule-derived labels, which are useful for engineering validation but do not replace longitudinal outcomes from real applicants. Future work should include consent-based user data, pre-test and post-test score change, click-through on recommended materials, acceptance of ranked recommendations, Precision@k and MRR for retrieval, citation correctness, and hallucination-rate comparison with and without retrieval.

X. Conclusion

This paper presented an AI-based assistant for selecting educational direction and preparing for exams according to student ability. The system combines topic diagnosis, adaptive mastery-aware blending, LambdaRank educational recommendation, rule-based explanations, and a local-source grounded assistant. The results indicate that a hybrid approach is practical for low-data educational settings where interpretability, regional constraints, and citation-backed answers are essential.

Acknowledgment

The authors thank the Department of Computer Engineering at Kyrgyz-Turkish Manas University for supervision and feedback during the development of the prototype.

References

[1] P. Brusilovsky and E. Millan, "User models for adaptive hypermedia and adaptive educational systems," in *The Adaptive Web*, Lecture Notes in Computer Science, vol. 4321. Springer, 2007, pp. 3–53.

[2] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein, "Deep knowledge tracing," in *Advances in Neural Information Processing Systems*, 2015, pp. 505–513.

[3] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User Modeling and User-Adapted Interaction*, vol. 4, no. 4, pp. 253–278, 1994.

[4] K. Verbert, N. Manouselis, X. Ochoa, M. Wolpers, H. Drachsler, I. Bosnic, and E. Duval, "Context-aware recommender systems for learning: A survey and future challenges," *IEEE Transactions on Learning Technologies*, vol. 5, no. 4, pp. 318–335, 2012.

[5] H. Nakagawa, Y. Iwasawa, and Y. Matsuo, "Graph-based knowledge tracing: Modeling student proficiency using graph neural networks," *Web Intelligence*, vol. 19, no. 1–2, pp. 87–102, 2021.

[6] A. Esteban, A. Zafra, and C. Romero, "Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization," *Knowledge-Based Systems*, vol. 194, p. 105385, 2020.

[7] D. K. Bokde, S. Girase, and D. Mukhopadhyay, "An approach to a university recommendation by multi-criteria collaborative filtering and dimensionality reduction techniques," in *Proceedings of the IEEE International Symposium on Nanoelectronic and Information Systems*, 2015, pp. 231–236.

[8] M. M. Khajah, R. M. Wing, R. V. Lindsey, and M. C. Mozer, "Integrating latent-factor and knowledge-tracing models to predict individual differences in learning," in *Proceedings of the 7th International Conference on Educational Data Mining*, 2014, pp. 99–106.

[9] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems*, 2020.

[10] C. J. C. Burges, "From RankNet to LambdaRank to LambdaMART: An overview," Microsoft Research, Tech. Rep. MSR-TR-2010-82, 2010.

[11] K. Jarvelin and J. Kekalainen, "Cumulated gain-based evaluation of IR techniques," *ACM Transactions on Information Systems*, vol. 20, no. 4, pp. 422–446, 2002.

[12] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019, pp. 8024–8035.

[13] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the International Conference on Machine Learning*, 2015, pp. 448–456.

[14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[15] S. Ramirez, "FastAPI," 2018. [Online]. Available: https://fastapi.tiangolo.com/

[16] V. Karpukhin *et al.*, "Dense passage retrieval for open-domain question answering," in *Proceedings of EMNLP*, 2020, pp. 6769–6781.

[17] S. Robertson and H. Zaragoza, "The probabilistic relevance framework: BM25 and beyond," *Foundations and Trends in Information Retrieval*, vol. 3, no. 4, pp. 333–389, 2009.

[18] H. Drachsler, H. G. K. Hummel, and R. Koper, "Personal recommender systems for learners in lifelong learning networks: The requirements, techniques and model," *International Journal of Learning Technology*, vol. 3, no. 4, pp. 404–423, 2008.

[19] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, 2005.

[20] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," in *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, 2009, pp. 452–461.

[21] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural collaborative filtering," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 173–182.

[22] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M.-W. Chang, "Retrieval augmented language model pre-training," in *Proceedings of the 37th International Conference on Machine Learning*, 2020, pp. 3929–3938.

DigitalClone: кыргыз-орус тилдүү чөйрөдө санариптик клон түзүү жана AI датасеттерин жыйноо үчүн HCI-платформа

Ислам Женишбеков
Компьютердик инженерия
Кыргыз-Түрк «Манас» университети
Бишкек, Кыргызстан
2104.01008@manas.edu.kg

Чинара Жумабаева
Компьютердик инженерия
Кыргыз-Түрк «Манас» университети
Бишкек, Кыргызстан
chinara.jumabaeva@manas.edu.kg

***Абстракт*—Кыргыз-орус тилдүү чөйрөдө адамдын санариптик клонун түзүү кыргыз тили үчүн сүйлөшүү стилиндеги тексттик маалыматтардын жана сапаттуу белгиленген аудио-текст датасеттеринин жетишсиздиги менен татаалдашат [2], [10], [11]. Бул чектөөлөр кеп таануу, кеп синтези жана диалогдук системаларды өнүктүрүүгө түз таасир этет. Ушул маселени чечүү үчүн адам-компьютер өз ара аракеттенүүсү (Human-Computer Interaction, HCI) [1] принциптерине негизделген DigitalClone платформасы сунушталат. Платформа колдонуучуга өзүнүн жазуу стилин, жеке маалыматын жана үн өзгөчөлүктөрүн эске алган санариптик клон түзүү мүмкүнчүлүгүн берет. Ошол эле учурда колдонуучунун макулдугу менен жазышуу архивдеринен алынган каттары жана үн үлгүлөрү тазаланып, структуралаштырылып, кыргыз тили үчүн табигый тексттик жана аудио датасеттерди түзүүгө негиз болот. Системанын архитектурасында чоң тил модели (Large Language Model, LLM), контекст менен толукталган издөө ыкмасы (Retrieval-Augmented Generation, RAG), автоматтык кеп таануу (Automatic Speech Recognition, ASR) жана тексттен кепке айландыруу (Text-to-Speech, TTS) модулдары бириктирилип, клон түзүү менен сапаттуу AI ресурстарын топтоо бир процесске бириктирилет.**

Ачкыч сөздөр—жасалма интеллект, датасет, кеп таануу, үн клондоо, RAG, LLM

I. КИРИШҮҮ

Акыркы жылдары чоң тилдик моделдер, автоматтык кеп таануу жана тексттен кепке айландыруу технологиялары адам менен компьютердин өз ара аракеттенүүсүн жаңы деңгээлге чыгарды. Санариптик системалар мурда негизинен меню, форма жана баскычтар аркылуу башкарылса, азыр колдонуучу табигый тилде суроо берип, жооп алып жана үн аркылуу баарлаша алган интеллектуалдык интерфейстер кеңири жайылууда. Бул өзгөрүү адамдын жазуу стилин, жеке маалыматын жана үн өзгөчөлүктөрүн эске алган санариптик клон системаларына кызыгууну күчөттү.

Кыргыз тили үчүн мындай системаларды түзүүдө негизги тоскоолдук — табигый сүйлөшүү стилиндеги тексттик маалыматтардын жана сапаттуу белгиленген аудио-текст датасеттеринин жетишсиздиги [2], [10], [11], [12]. Ачык булактардагы кыргызча материалдардын көпчүлүгү расмий, жаңылыктык же адабий стилге жакын, ал эми күнүмдүк жазышуулар, кыргыз-орус аралаш тексттер жана үн клондоо үчүн ылайыктуу аудио материалдар чектелүү. Бул жагдай кеп таануу, кеп синтези жана диалогдук системалардын сапатына түз таасир этет.

Ушул көйгөйдү чечүү үчүн DigitalClone платформасы сунушталат. Платформа колдонуучуга өзүнүн жазуу стилин, жеке контекстин жана үн өзгөчөлүктөрүн эске алган санариптик клон түзүүгө мүмкүндүк берет. Ошол эле учурда колдонуучунун макулдугу менен жазышуу архивдеринен жана үн үлгүлөрүнөн алынган маалыматтар тазаланып, структуралаштырылып, кыргыз тили үчүн табигый тексттик жана аудио датасеттерди түзүүгө негиз боло алат.

II. КЫРГЫЗ ТИЛИНДЕГИ AI РЕСУРСТАРЫНЫН ЖЕТИШСИЗДИГИ

Кыргыз тили үчүн AI ресурстары акыркы жылдары көбөйүп жатканы менен, санариптик клон, кеп таануу жана кеп синтези сыяктуу системалар үчүн алар дагы эле жетишсиз [2], [10], [11], [12]. Бул жетишсиздик эки негизги багытта байкалат: биринчиси — тексттик корпустардын тематикасы жана стили чектелүү: жаңылыктар, адабият ж.б., ал эми табигый, күнүмдүк жашоодон келген кептер жок, экинчиси — сапаттуу белгиленген аудио-текст датасеттеринин көлөмү аз.

ЖАДЫБАЛ I
КЫРГЫЗ ТИЛИНДЕГИ НЕГИЗГИ ТЕКСТТИК ДАТАСЕТТЕР ЖАНА АЛАРДЫН ТЕМАТИКАСЫ

Датасет	Негизги тематикасы жана мүнөзү
KyrText [2]	жаңылыктар, Wikipedia макалалары, адабият ж.б.
Kyrgyz Wikipedia [6]	энциклопедиялык макалалар, терминдер
CC100 [7] / OSCAR [8] / mC4 [9]	веб-баракчалардан алынган жалпы тексттер, жаңылыктар
Kyrgyz News Corpus [3]	жаңылыктар
Manas-UdS Kyrgyz Corpus [4]	көркөм адабият, эпос, жомок, гезиттердин тексттери
UD_Kyrgyz-KTMU [5]	синтаксистик белгиленген кыргызча тексттер

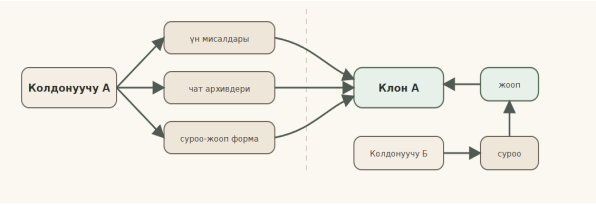
Жадыбал II

ТИЛДЕР БОЮНЧА АЧЫК АУДИО-ТЕКСТ ДАТАСЕТТЕРИНИН КӨЛӨМҮН САЛЫШТЫРУУ

Тил	Болжолдуу көлөмү, саат
Англис [16], [17]	50000+
Орус [15]	20000+
Казах [13], [14]	1200+
Кыргыз [10], [11], [12]	~70

III. ПЛАТФОРМАНЫН ЖАЛПЫ КОНЦЕПЦИЯСЫ

Кыргыз тили үчүн табигый тексттик жана үндүк датасеттерди түзүүдөгү негизги тоскоолдук — мындай маалыматтардын көбү жеке мүнөздө болушу. Күнүмдүк кат алышуулар, үн билдирүүлөрү жана сүйлөшүү үлгүлөрү ачык интернет булактарында жетиштүү көлөмдө кездешпейт жана аларды колдонуучунун уруксатысыз чогултуу этикалык жана укуктук жактан туура эмес [29]. Ошондуктан бул иште маалымат жыйноо процесси колдонуучу үчүн реалдуу пайдалуу кызмат менен байланыштырылат: адам өзүнүн санариптик клонун түзөт, ал эми система анын макулдугу менен таза, структураланган жана табигый тилдик материалдарды даярдайт, бирок ошол эле учурда сөзсүз купуялыгын сактайт. DigitalClone платформасынын негизги идеясы ушул эки тараптуу баалуулукка негизделет.



Сүрөт I. Клон түзүү жана клон менен баарлашуу концепциясы

IV. БЕРИЛИШТЕРДИ КАБЫЛ АЛУУ ЖАНА ДАЯРДОО

DigitalClone платформасында санариптик клон түзүү үчүн маалыматтар үч негизги булактан алынат: үн үлгүлөрү, суроо-жооп формасы жана чат архивдери. Бул булактар бири-бирин толуктайт: үн үлгүлөрү клондун сүйлөө өзгөчөлүгүн берсе, форма туруктуу жеке контекстти берет, ал эми чат архивдери адамдын табигый жазуу стилин жана күнүмдүк баарлашуу манерасын көрсөтөт. Маалыматтар түздөн-түз моделге берилбестен, алдын ала тазаланып, структуралаштырылып жана кийин колдонууга ыңгайлуу форматка өткөрүлөт.

А. Үн үлгүлөрү

Үн клондоо жана TTS үчүн колдонуучуга атайын сүйлөмдөр топтому берилет. Бул сүйлөмдөр кыска жана узун фразаларды, суроо сүйлөмдөрдү, эмоцияны билдирген репликаларды, сан жана аталыштарды, ошондой эле кыргыз-орус чөйрөсүндө көп кездешкен сөздөрдү камтышы керек. Мындай ар түрдүүлүк моделге адамдын үн тембрин гана эмес, интонациясын жана сүйлөө ритмин да жакшыраак үйрөнүүгө жардам берет.

Колдонуучу бул тексттерди окуп үн жазат. Андан кийин аудио сегменттерге бөлүнүп, ар бир сегмент тиешелүү текст менен дал келтирилет. Натыйжада TTS үчүн колдонууга ыңгайлуу аудио-текст жуптары түзүлөт. Бул ыкма жөн гана узун аудио жүктөөдөн жакшыраак, анткени модель үчүн таза, кыска жана так белгиленген материал алынат.

Б. Суроо-жооп формасы

Форма клондун негизги жеке контекстин берет. Анда колдонуучунун өзү тууралуу маалыматтары, сүйлөө стили, кызыкчылыктары жана клон кандай тон менен жооп бериши керектиги такталат. Бул маалымат чат архивдеринен автоматтык түрдө дайыма эле түшүнүктүү боло бербейт, ошондуктан форма клон үчүн туруктуу багыт катары колдонулат.

Форманын артыкчылыгы — маалымат алдын ала структураланган түрдө алынат. Мисалы, “кандай тилде көп сүйлөйт?”, “жооптору расмийби же бейформалбы?”, “клон кайсы ролдо сүйлөшү керек?” деген суроолор кийин жооп түзүүдө түздөн-түз контекст катары колдонулушу мүмкүн.

В. Чат архивдери

Чат архивдери колдонуучунун табигый жазуу стилин алуу үчүн колдонулат. Архив жүктөлгөндөн кийин система файлдын түрүн аныктап, билдирүүлөрдү бөлүп чыгарат. Андан соң колдонуучу аты (username) же башка идентификатор аркылуу керектүү колдонуучунун гана билдирүүлөрү тандалат.

Тандалган билдирүүлөр тазаланат: кайталанган, өтө кыска же техникалык мааниси жок фрагменттер алынып салынат. Ар бир билдирүүгө текст, тил, убакыт, булак жана чаттын ролу сыяктуу метадайындар кошулат. Натыйжада жеке тексттик корпус түзүлөт.

Андан кийин бул корпус маанилик чанктарга бөлүнөт. Ар бир чанк эмбеддинг-вектор (embedding vector) түрүндө сакталат, ал эми баштапкы текст жана метадайындар өзүнчө байланышта калат. Бул структура кийинки этапта суроого жакын билдирүүлөрдү тез табууга мүмкүндүк берет.

V. ЖООП ТҮЗҮҮ ЧЫНЖЫРЫ

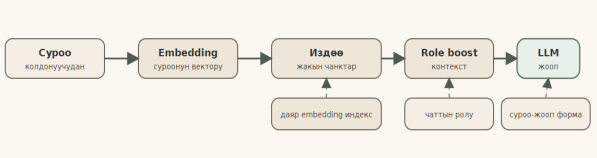
Жооп түзүү этабында система даярдалган маалыматтарды колдонот. Негизги максат — LLMге мүмкүн болушунча көп маалымат берүү эмес, суроого тиешелүү жана контекстке жакын маалыматты тандоо.

Алгач колдонуучунун суроосу кабыл алынат. Эгер суроо үн аркылуу берилсе, ал ASR модулу аркылуу текстке айландырылат. Андан кийин суроонун эмбеддинг-вектору түзүлүп, алдын ала даярдалган жеке корпус индекси менен салыштырылат.

Smart RAG механизми Retrieval-Augmented Generation ыкмасына таянып [18], суроого семантикалык жактан жакын чанктарды табат. Эгер чаттын ролу суроонун контекстине туура келсе, мисалы дос, үй-бүлө же жумуш чөйрөсү, ошол чанктарга кошумча салмак берилет. Натыйжада LLM туш келди мисалдарды эмес, суроого жана кырдаалга жакын болгон тексттик контекстти алат.

Жооп түзүүдө LLM үч негизги маалыматты колдонот: суроонун өзүн, форма аркылуу алынган туруктуу жеке контекстти жана Smart RAG аркылуу табылган релеванттуу билдирүүлөрдү. Бул ыкма клондун жообун

жалпы AI жардамчынын жообунан айырмалап, аны колдонуучунун стилине жана мурдагы баарлашуу манерасына жакындатат.



Сүрөт II. Жооп түзүү чынжыры

VI. ASR ЖАНА TTS

Бул бөлүмдө платформанын үндүк бөлүгү эки өзүнчө тапшырма катары каралат: ASR колдонуучунун үнүн текстке айлантат, ал эми TTS клондун тексттик жообун үнгө айландырат. Максат — даяр API кызматтарын жана локалдуу иштетиле турган моделдерди салыштырып, кыргыз тили үчүн кайсы ыкмалар практикалык жактан ылайыктуу экенин баалоо.

A. ASR

ASR сапатын баалоо үчүн кыргызча аудио үлгүлөрдөн турган тесттик топтом даярдалды. Ага FLEURS [11], Common Voice [10] жана кошумча автордук мисалдар киргизилди. Кыргыз жана орус тилдери менен иштеген көп тилдүү ASR моделдеринин орус тили боюнча ачык эталондук баалоолорунда WER 10%дан төмөн деңгээлде берилген [30]. Ошондуктан бул иште 10% чеги кыргыз тилиндеги натыйжаларды салыштыруу үчүн багыттоочу ориентир катары колдонулду. Ар бир модель бир эле аудио файлдарды транскрипциялайт, андан кийин натыйжа эталон текст менен салыштырылып, сөз катасынын көрсөткүчү (Word Error Rate, WER) эсептелет. WER канчалык төмөн болсо, модель кепти ошончолук так тааныган болуп эсептелет. Салыштыруу 3-жадыбалда көрсөтүлгөн.

Б. TTS

TTS бөлүгүндө жөн гана текстти үнгө айландыруу эмес, конкреттүү колдонуучунун үнүн клондоо мүмкүнчүлүгү салыштырылат. Бул үчүн даяр API кызматы катары ElevenLabs [24] жана автор тарабынан түзүлгөн локалдуу TTS пайплайндары каралат. TTS сапатын баалоо үчүн MOS (Mean Opinion Score) ыкмасы колдонулат [28]. Баалоочулар ар бир аудиону 1ден 5ке чейинки шкалада табигый угулушу жана оригиналдуу үнгө окшоштугу боюнча баалашат. Баа канчалык жогору болсо, синтезделген үн ошончолук табигый жана оригиналдуу үнгө жакын деп эсептелет. Тест үчүн кыргызча жана орусча сүйлөмдөрдүн бирдей топтому даярдалып, ар бир TTS системасы ошол тексттерди үнгө айландыруу үчүн колдонулат. Салыштыруу 4-жадыбалда көрсөтүлгөн.

Жадыбал III

ASR моделдерин кыргыз тилиндеги WER боюнча салыштыруу

Модель	WER, %	Локалдуу	Кыргыз, орус тилдүү
AIrun Standard [21]	2.21	жок	жок
Gemini 3.5 Flash [19]	7.28	жок	ооба
Gemini 2.5 Pro [19]	9.05	жок	ооба
GPT-4o Transcribe [20]	20.53	жок	ооба
UlutSoftLLC/whisper-small-kyrgyz [22]	26.49	ооба	жок
nineninesix/kyrgyz-whisper-medium [23]	30.24	ооба	жок

Жадыбал IV

Үн клондоо системаларын MOS боюнча салыштыруу

Модель	MOS	Локалдуу модель
ElevenLabs [24] орусча Professional Cloning	3.5	жок
ElevenLabs [24] орусча Instant Cloning	3	жок
ElevenLabs [24] кыргызча Professional Cloning	2.3	жок
XTTS [26] + RVC [27] орусча локалдык пайплайны (автордук 20 мин. датасетте үйрөтүү)	1.5	ооба
KaniTTS [25] + LoRA кыргызча локалдык пайплайны (автордук 5 сааттык датасетте үйрөтүү)	1.4	ооба

VII. КУПУЯЛУУЛУК

Санариптик клон системасы жеке чаттар, суроо-жооп формасы жана үн үлгүлөрү менен иштегендиктен, купуялуулук негизги талаптардын бири болуп саналат [29]. Бул маалыматтарда колдонуучунун жеке ойлору, байланыштары, сүйлөө стили, үнү жана үчүнчү тараптарга тиешелүү маалыматтар болушу мүмкүн. Ошондуктан мындай материалдар жөнөкөй техникалык ресурс эмес, сезимтал жеке маалымат катары каралышы керек.

ASR жана TTS боюнча жүргүзүлгөн салыштыруулар учурда айрым API кызматтары локалдуу моделдерге караганда жакшыраак сапат бере аларын көрсөтөт. Практикалык система үчүн бул маанилүү, анткени кепти туура таануу жана үндү табигый синтездөө клондун сапатына түз таасир этет. Бирок API колдонуу маалыматты тышкы кызматка жөнөтүүнү талап кылат. Жеке чаттар жана үн үлгүлөрү менен иштеген платформада бул олуттуу купуялуулук тобокелдигин жаратат.

Ушул себептен узак мөөнөттүү багыт катары ASR, TTS жана LLM моделдерин мүмкүн болушунча локалдуу иштетүүгө умтулуу зарыл. Локалдуу моделдер сапат жагынан API кызматтарына толук жетпесе да, алар колдонуучунун маалыматтарын сыртка чыгарбай иштетүүгө мүмкүндүк берет. Демек, платформада сапат менен купуялуулуктун ортосунда тең салмак табуу керек: учурда жакшы натыйжа берген API моделдер колдонулушу мүмкүн, бирок изилдөөнүн негизги

багыттарынын бири локалдуу моделдердин сапатын жогорулатуу болушу керек.

Маалыматтарды сактоо жана иштетүүдө да этияттык талап кылынат. Колдонуучу жүктөгөн архивдер, үн жазуулары жана түзүлгөн корпустар корголгон чөйрөдө сакталууга тийиш. Маалыматтарга жетүү укугу чектелип, керексиз файлдар өчүрүлүп, датасетке кирбей турган жеке же техникалык фрагменттер алдын ала тазаланышы керек. Айрыкча үн үлгүлөрү биометрикалык маалымат катары каралышы мүмкүн болгондуктан, аларды уруксатсыз көчүрүү же үчүнчү тарапка берүү туура эмес.

Эгер платформа аркылуу алынган материалдар кийин ачык датасет катары жарыяланса, алар сөзсүз түрдө алдын ала анонимдештирүү процессинен өтүшү керек. Жеке ысымдар, телефон номерлери, даректер, аккаунттар, окуу же жумуш орду сыяктуу идентификациялоочу маалыматтар алынып салынышы зарыл. Ошондо гана мындай маалыматтарды кыргыз тили үчүн AI ресурстарын өнүктүрүүдө этикалык жана коопсуз колдонууга болот.

VIII. КОРУТУНДУ

Бул иш кыргыз тили үчүн санариптик клон түзүүнү жеке AI кызматы катары гана эмес, табигый тексттик жана үндүк датасеттерди топтоонун практикалык жолу катары да карады. DigitalClone платформасы колдонуучу үчүн санариптик клон түзүү мүмкүнчүлүгүн жана кыргыз тили үчүн сапаттуу маалымат топтоо идеясын бириктирет.

Иштин алкагында колдонуучунун билдирүүлөрүнө жакын жооп түзүү үчүн Smart RAG логикасы сунушталды, кыргыз тилиндеги ASR жана үн клондоо чечимдери салыштырылды, ошондой эле кыргыз жана орус тилдери үчүн локалдуу үн клондоо пайплайндары иштелди. Кошумча салым катары кыргыз тили үчүн автордук үн материалы топтолду. Мындан ары негизги багыттар — локалдуу моделдердин сапатын жогорулатуу, маалыматтарды коопсуз анонимдештирүү жана кыргыз тили үчүн ачык, сапаттуу датасеттерди кеңейтүү болуп саналат.

ПАЙДАЛАНЫЛГАН БУЛАКТАР

[1] NIST. Human Factors: Human-Centered Design. https://www.nist.gov/itl/iad/visualization-and-usability-group/human-factors-human-centered-design

[2] M. S. Mambetisaeva et al. KyrText: A Corpus for Kyrgyz Language Modeling. https://aclanthology.org/2026.loreslm-1.39/

[3] The Cramer Project. Kyrgyz News Corpus. https://huggingface.co/datasets/the-cramer-project/Kyrgyz_News_Corpus

[4] Manas-UdS Kyrgyz Corpus. https://fedora.clarin-d.uni-saarland.de/kyrgyz/index.html

[5] Universal Dependencies. UD_Kyrgyz-KTMU. https://universaldependencies.org/treebanks/ky_ktmu/

[6] Wikimedia Downloads. Kyrgyz Wikipedia dumps. https://dumps.wikimedia.org/kywiki/

[7] A. Conneau et al. Unsupervised Cross-lingual Representation Learning at Scale. https://aclanthology.org/2020.acl-main.747/

[8] OSCAR Project. Open Super-large Crawled Aggregated coRpus. https://oscar-project.org/

[9] TensorFlow Datasets. C4 / mC4 Dataset. https://tensorflow.google.cn/datasets/catalog/c4

[10] Mozilla Common Voice. Kyrgyz dataset. https://prod.datacollective.mozillafoundation.org/datasets/cmflnuzw6wo7l6j5jn70lwwcn

[11] Google Research. FLEURS dataset. https://huggingface.co/datasets/google/fleurs

[12] Simonlob. Kany dataset. https://huggingface.co/datasets/Simonlob/Kany_dataset_mk4_Base

[13] A. Khassanov et al. A Crowdsourced Open-Source Kazakh Speech Corpus and Initial Speech Recognition Baseline. https://arxiv.org/abs/2009.10334

[14] H. Mussakhojayeva et al. KSC2: An Industrial-Scale Open-Source Kazakh Speech Corpus. https://www.isca-archive.org/interspeech_2022/mussakhojayeva22_interspeech.pdf

[15] Microsoft Azure Open Datasets. Russian Open Speech To Text. https://learn.microsoft.com/en-us/azure/open-datasets/dataset-open-speech-text

[16] V. Pratap et al. Multilingual LibriSpeech. https://arxiv.org/abs/2012.03411

[17] D. Galvez et al. The People's Speech: A Large-Scale Diverse English Speech Recognition Dataset. https://arxiv.org/abs/2111.09344

[18] P. Lewis et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. https://arxiv.org/abs/2005.11401

[19] Google AI for Developers. Gemini API audio understanding. https://ai.google.dev/gemini-api/docs/audio

[20] OpenAI. GPT-4o Transcribe model documentation. https://platform.openai.com/docs/models/gpt-4o-transcribe

[21] Kyrgyz national news agency Kabar. First digital avatar in Kyrgyz language. https://en.kabar.kg/news/first-digital-avatar-in-kyrgyz-language-ai-is-not-privilege-but-national-achievement/

[22] UlutSoftLLC. whisper-small-kyrgyz. https://huggingface.co/UlutSoftLLC/whisper-small-kyrgyz

[23] nineninesix. kyrgyz-whisper-medium. https://huggingface.co/nineninesix/kyrgyz-whisper-medium

[24] ElevenLabs. Voice Cloning documentation. https://elevenlabs.io/docs/creative-platform/voices/voice-cloning

[25] nineninesix. KaniTTS Kyrgyz model. https://huggingface.co/nineninesix/kani-tts-400m-ky-kani

[26] Coqui. XTTS-v2. https://huggingface.co/coqui/XTTS-v2

[27] RVC Project. Retrieval-based Voice Conversion. https://github.com/RVC-Project/Retrieval-based-Voice-Conversion

[28] ITU-T. Recommendation P.800.1: Mean Opinion Score terminology. https://www.itu.int/rec/T-REC-P.800.1

[29] OECD. AI Principles. https://www.oecd.org/en/topics/ai-principles.html

[30] OpenAI. Whisper: robust speech recognition and WER breakdown by language. https://github.com/openai/whisper

ЖАСАЛМА ИНТЕЛЛЕКТ ЖАНА КРИТИКАЛЫК ОЙ ЖҮГҮРТҮҮ: БИЛИМ БЕРҮҮДӨГҮ ФИЛОСОФИЯЛЫК ЖАНА ПЕДАГОГИКАЛЫК ӨЗГӨРҮҮЛӨР

DIGITAL ARCHITECTURE AND SUSTAINABLE DEVELOPMENT: THE ROLE OF ARTIFICIAL INTELLIGENCE IN ENERGY-EFFICIENT DESIGN

Жаңыл Барпыбаева
И. Арабаев атындагы КМУ
Бишкек, Кыргызстан
janylbarpy@gmail.com

Zhanyl Barpybaeva
Kyrgyz State University named after I.arabaev
Bishkek, Kyrgyzstan
janylbarpy@gmail.com

Аннотация: Бул макала санариптик трансформация шартында жасалма интеллекттин билим берүү процессине тийгизген таасирин жана анын критикалык ой жүгүртүү менен өз ара байланышын философиялык жана педагогикалык жактан талдоого арналат. Изилдөөдө жасалма интеллекттин маалыматты жеткирүү жана иштетүү мүмкүнчүлүктөрү студенттердин ой жүгүртүү маданиятына кандай таасир этери каралат. Өзгөчө көңүл гуманитардык багыттагы liberal arts билим берүү моделинде кездешкен өзгөрүүлөргө бурулат, анткени бул система критикалык ой жүгүртүүнү өнүктүрүүгө негизделген. Макалада жасалма интеллект бир жагынан билим алууну жеңилдетүүчү инструмент катары, экинчи жагынан интеллектуалдык пассивдүүлүктү жараткан фактор катары каралат. Изилдөөнүн жыйынтыгында технологиянын таасири анын колдонулуу маданиятына түздөн-түз көз каранды экени аныкталат.

Ачкычсөздөр; жасалма интеллект, критикалык ой жүгүртүү, билим берүү, санариптик трансформация, коомдук-гуманитардык илимдер, философия, педагогика)

I. КИРИШҮҮ

Акыркы он жылдыкта дүйнөлүк коомчулук терең санариптик трансформация дооруна кадам басты. Маалыматтык технологиялардын ылдам өнүгүшү, өзгөчө жасалма интеллект системаларынын кеңири жайылышы билим берүү тармагына сапаттык жаңы өзгөрүүлөрдү алып келди. Бүгүнкү күндө билим берүү мейкиндиги салттуу окутуу моделдеринен чыгып, гибриддик жана санариптик форматтарга өтүп, билимдин жаралышы, таралышы жана өздөштүрүлүшү принципиалдуу түрдө өзгөрүүдө. Мындай шартта билим берүүнүн мазмуну гана эмес, анын философиялык негиздери да кайра кароону талап кылууда.

Жасалма интеллекттин билим берүү системасына активдүү кириши, айрыкча текст түзүүчү жана анализ жүргүзүүчү санариптик инструменттердин пайда болушу студенттердин окуу жүрүм-турумуна олуттуу таасир этүүдө. Бул көрүнүш, бир жагынан, билим алуунун жеткиликтүүлүгүн жогорулатып, маалыматты иштетүүнү жеңилдетсе, экинчи жагынан, ой жүгүртүүнүн табиятына байланыштуу жаңы суроолорду жаратат. Студент даяр маалыматты тез алуу мүмкүнчүлүгүнө ээ болгондо, анын өз алдынча ой жүгүртүү жөндөмү кандай өзгөрөт деген маселе бүгүнкү күндө өзгөчө актуалдуу болуп турат.

Айрыкча гуманитардык багыттагы, анын ичинде liberal arts билим берүү моделинде бул көйгөй тереңирээк көрүнөт. Анткени мындай билим берүү системасы студенттин критикалык ой жүгүртүүсүн, аргументация жөндөмүн, текстти интерпретациялоо жана өз алдынча позиция калыптандыруу көндүмдөрүн өнүктүрүүгө багытталган. Бирок жасалма интеллект инструменттеринин кеңири колдонулушу бул процессти жөнөкөйлөтүп гана тим болбостон, айрым учурларда формалдаштырып, ой жүгүртүүнүн ички процессин сырткы технологиялык механизмдер менен алмаштыруу коркунучун жаратат.

Бул контекстте критикалык ой жүгүртүү түшүнүгү өзгөчө мааниге ээ болууда. Критикалык ой жүгүртүү — бул маалыматты кабыл алуу эмес, аны талдоо, салыштыруу, күмөн саноо жана негиздүү жыйынтык чыгаруу жөндөмү катары каралат. Билим берүү системасы дал ушул жөндөмдү калыптандырууга багытталууга тийиш. Ошол эле учурда, жасалма интеллект бул процесске кандайча таасир этет деген суроо бир жактуу жоопко ээ эмес. Айрым изилдөөлөрдө жасалма интеллект студенттердин аналитикалык мүмкүнчүлүктөрүн кеңейтет деп белгиленсе [1, 45-б.], башкаларында тескерисинче когнитивдик активдүүлүктүн төмөндөшүнө алып келиши мүмкүн экени көрсөтүлөт [2, 112-б.].

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Мындай карама-каршылыктуу баалоолор изилдөөнүн предметин жана актуалдуулугун аныктайт. Бул макалада жасалма интеллект менен критикалык ой жүгүртүүнүн өз ара байланышы билим берүү философиясынын алкагында каралат. Изилдөөнүн объектиси катары жогорку билим берүү системасындагы окуу процесси алынат, ал эми предмети — жасалма интеллект технологияларынын студенттердин критикалык ой жүгүртүү жөндөмүнө тийгизген таасири болуп саналат.

Изилдөөнүн негизги максаты — санариптик трансформация шартында жасалма интеллекттин билим берүү процессиндеги ордун аныктоо менен анын критикалык ой жүгүртүүгө болгон таасирин философиялык жана педагогикалык жактан талдоо болуп саналат. Бул максатка жетүү үчүн төмөнкү милдеттер коюлат: критикалык ой жүгүртүү түшүнүгүнүн теориялык негиздерин ачып көрсөтүү; жасалма интеллекттин билим берүүдө колдонулуш өзгөчөлүктөрүн талдоо; анын оң жана терс таасирлерин аныктоо; ошондой эле практикалык байкоолордун негизинде билим берүү процессинде пайда болгон өзгөрүүлөрдү баалоо.

Методологиялык негиз катары философиялык анализ, салыштырма ыкма, ошондой эле педагогикалык байкоо ыкмалары колдонулат. Изилдөөнүн эмпирикалык базасы катары автордун окутуучулук тажрыйбасында байкалган көрүнүштөр, студенттердин окуу процессиндеги жүрүм-турум өзгөрүүлөрү жана жасалма интеллект инструменттерин колдонуу практикасы алынат.

Бул изилдөөнүн илимий жаңылыгы жасалма интеллектти билим берүүдө жөн гана технологиялык инструмент катары эмес, ой жүгүртүүнүн структурасына таасир этүүчү фактор катары кароодо көрүнөт. Ошондой эле, гуманитардык билим берүү контекстинде, айрыкча liberal arts моделинде, критикалык ой жүгүртүүнүн трансформациясын талдоого өзгөчө көңүл бурулат. f

II. ТЕОРИЯЛЫК НЕГИЗДЕР

III. КРИТИКАЛЫК ОЙ ЖҮГҮРТҮҮНҮН ФИЛОСОФИЯЛЫК МАҢЫЗЫ

Критикалык ой жүгүртүү түшүнүгү философия тарыхында өзгөчө орунду ээлейт жана ал жөн гана логикалык операциялардын жыйындысы эмес, адамдын дүйнөнү аңдоо ыкмасы катары каралат. Ой жүгүртүүнүн бул формасы чындыкты даяр кабыл алуудан баш тартып, ар бир маалыматты күмөн саноо, талдоо жана негиздөө аркылуу кабыл алууну талап кылат. Мындай мамиле байыркы философиядан тартып бүгүнкү күнгө чейин интеллектуалдык маданияттын негизги белгиси катары бааланып келген.

Критикалык ой жүгүртүүнүн башаттарын антикалык философиядан табууга болот. Socrates өзүнүн диалогдук ыкмасы аркылуу чындыкка жетүүнүн жолу суроо берүү жана күмөн саноо аркылуу ишке ашарын көрсөткөн. Анын методунда жооптон көрө туура коюлган суроо маанилүүрөөк болуп, бул бүгүнкү күндөгү критикалык ой жүгүртүүнүн өзөгүн түзөт. Сократтык ыкмада адам өзүнүн билбестигин аңдап, ошол аркылуу билимге умтулат. Бул принцип заманбап билим берүүдө да өз актуалдуулугун жоготкон жок.

Жаңы доор философиясында критикалык ой жүгүртүү рационалдуулук менен тыгыз байланышта каралган. René Descartes «методикалык күмөн саноо» принцибин сунуштап, ар бир билимди текшерүүнү талап кылган [3, 58-б.]. Анын ою боюнча, чыныгы билим күмөн саноодон өткөн гана учурда ишенимдүү болуп саналат. Бул ыкма ой жүгүртүүнү активдүү процесс катары түшүнүүгө өбөлгө түзүп, адамды пассивдүү кабыл алуудан алыстатат.

Андан ары Immanuel Kant критикалык философиянын негизин түзүп, ой жүгүртүүнүн чектерин жана мүмкүнчүлүктөрүн изилдеген. Анын концепциясында «критика» түшүнүгү четке кагуу эмес, тескерисинче акылдын мүмкүнчүлүктөрүн аңдап түшүнүү катары берилет [4, 112-б.]. Канттын бул идеясы критикалык ой жүгүртүүнү деструктивдүү эмес, конструктивдүү интеллектуалдык ишмердүүлүк катары кароого мүмкүндүк берет.

Заманбап философия жана педагогикада критикалык ой жүгүртүү көбүнчө рефлексия, аргументация жана интерпретация менен байланышта каралат. John Dewey критикалык ой жүгүртүүнү «рефлексивдүү ой жүгүртүү» катары

аныктап, аны тажрыйба менен байланышта өнүгүүчү процесс деп эсептеген [5, 76-б.]. Анын пикиринде, билим берүү студентти даяр жооптор менен камсыз кылууга эмес, маселени өз алдынча чечүүгө үйрөтүүгө тийиш.

Бул философиялык мамилелердин баары бир маанилүү ойго алып келет: критикалык ой жүгүртүү — бул билимди топтоо эмес, билим менен иштөө маданияты. Ал маалыматты жөн гана кабыл алуу эмес, аны кайра иштеп чыгуу, кайра карап чыгуу жана өз позициясын калыптандыруу процессин камтыйт. Ушул контекстте ой жүгүртүү адамдын интеллектуалдык автономиясынын негизги көрсөткүчү болуп саналат.

Бирок санариптик доордо бул түшүнүк жаңы сыноолорго туш болууда. Маалыматтын ашыкча көлөмү жана аны иштетүүнүн автоматташтырылышы ой жүгүртүүнүн ролун өзгөртүп жатат. Эгер мурда билим алуу маалыматты табуу менен байланыштуу болсо, бүгүнкү күндө негизги маселе — маалыматты туура тандоо жана интерпретациялоо болуп калды. Бул жагдай критикалык ой жүгүртүүнүн маанисин дагы да күчөтөт.

Ошол эле учурда, даяр маалыматка жеткиликтүүлүктүн жогорулашы адамдын өз алдынча ой жүгүртүү аракетин азайтышы мүмкүн деген коркунуч да бар. Айрым учурларда студент маалыматты терең талдоонун ордуна, аны даяр формада кабыл алууга көнүп калууда. Бул көрүнүш критикалык ой жүгүртүүнүн формалдуу мүнөзгө өтүп кетишине алып келиши мүмкүн, башкача айтканда, ой жүгүртүү процесси сакталгандай көрүнгөнү менен, анын мазмуну үстүртөн болуп калат.

Дал ушул себептен улам критикалык ой жүгүртүү бүгүнкү күндө жөн гана когнитивдик көндүм эмес, философиялык жана маданий феномен катары каралууда. Ал адамдын маалыматтык чөйрөдө өз ордун табуусуна, манипуляциядан коргонуусуна жана интеллектуалдык эркиндикти сактоосуна шарт түзөт. Билим берүү системасы бул сапатты калыптандырууну негизги максаттарынын бири катары карашы керек.

Жасалма интеллект түшүнүгү акыркы жылдары илимий гана эмес, коомдук дискурстун да борборуна айланды. Алгач тар чөйрөдө техникалык термин катары колдонулуп келген бул түшүнүк бүгүнкү күндө билим берүү, экономика, маданият жана күнүмдүк турмуш менен тыгыз байланышкан көп кырдуу феноменге айланды. Жалпысынан жасалма интеллект деп адамдын интеллектуалдык функцияларын — үйрөнүү, логикалык ой жүгүртүү, чечим кабыл алуу жана тилди иштетүү сыяктуу процесстерди аткара алган алгоритмдер жана системалар түшүнүлөт [6, 27-б.].

Жасалма интеллекттин өнүгүшү өзгөчө машиналык окутуу жана нейрондук тармактардын пайда болушу менен жаңы деңгээлге көтөрүлдү. Бул технологиялар чоң көлөмдөгү маалыматтарды талдоо аркылуу мыйзам ченемдүүлүктөрдү табууга жана ошонун негизинде алдын ала жыйынтык чыгарууга мүмкүндүк берет. Мындай системалар классикалык программалоодон айырмаланып, алдын ала так жазылган эрежелер менен гана чектелбестен, өз алдынча «үйрөнүү» жөндөмүнө ээ болуп калды [7, 91-б.].

Билим берүү тармагында жасалма интеллекттин колдонулушу ар кандай формада байкалууда. Бир жагынан, ал административдик жана уюштуруучулук процесстерди оптималдаштыруу үчүн колдонулса, экинчи жагынан түздөн-түз окутуу процессине интеграцияланып жатат. Мисалы, автоматташтырылган баалоо системалары, адаптивдүү окутуу платформалары жана текст түзүүчү интеллектуалдык инструменттер студенттердин билим алуу тажрыйбасын өзгөртүп жатат. Мындай инструменттер студентке керектүү маалыматты тез табууга, түшүндүрүүгө жана структуралоого жардам берет.

Айрыкча табигый тилди иштетүү технологияларынын өнүгүшү билим берүүдө сапаттык өзгөрүүлөрдү жаратты. Бул багыттагы системалар адам тилин түшүнүп, ага жооп кайтарууга жөндөмдүү болуп, студент менен «диалог жүргүзө» алган деңгээлге жетти. Мындай мүмкүнчүлүк билим алууну интерактивдүү кылып, студенттин суроо берүү жана жооп алуу процессин жеңилдетет. Ошол эле учурда бул көрүнүш билим менен ой жүгүртүүнүн ортосундагы чек араны кайра карап чыгууга мажбурлайт.

Бүгүнкү күндө кеңири колдонулуп жаткан ChatGPT сыяктуу системалар текст түзүү, түшүндүрүү жана талдоо функцияларын аткаруу менен студенттер үчүн универсалдуу жардамчы куралга айланып баратат. Алар кыска убакытта чоң көлөмдөгү маалыматты иштеп чыгып, логикалык жактан түзүлгөн жоопторду сунуштай алат. Бул, албетте, билим алуунун

Identify applicable funding agency here. If none, delete this text box.

эффективдүүлүгүн жогорулатат. Бирок ошол эле учурда мындай системалардын жооптору студенттин өз алдынча ой жүгүртүү процессин канчалык деңгээлде алмаштырары тууралуу суроо жаралат.

Жасалма интеллекттин билим берүүдө ролун бир жактуу баалоо мүмкүн эмес. Ал бир эле учурда мүмкүнчүлүк да, чакырык да болуп саналат. Мүмкүнчүлүк катары ал билим алууну демократиялаштырып, маалыматка жеткиликтүүлүктү кеңейтет. Мурда жеткиликсиз болгон билим булактары бүгүн ар бир студент үчүн ачык болуп калды. Мындан тышкары, жасалма интеллект индивидуалдуу окутууну ишке ашырууга шарт түзөт, башкача айтканда, ар бир студенттин деңгээлине жана муктаждыгына жараша билим берүү мүмкүнчүлүгү пайда болот.

Бирок чакырык катары караганда, жасалма интеллект билим берүүнүн маңызын өзгөртүү коркунучун да камтыйт. Эгер билим берүү процесси даяр жоопторду алуу менен гана чектелип калса, анда билимдин тереңдиги жана сапаты төмөндөшү мүмкүн. Мындай шартта студент маалыматты өз алдынча изилдөөнүн ордуна, даяр интерпретацияларды кабыл алууга көнүп калат. Бул көрүнүш, өзгөчө гуманитардык билим берүүдө, олуттуу көйгөй жаратышы мүмкүн, анткени бул багытта билим алуу процесси көбүнчө интерпретация, талдоо жана аргументацияга негизделет.

Санариптик трансформация шартында билим берүү системасы жаңы парадигмага өтүп жатат. Бул парадигмада мугалимдин ролу да өзгөрүүдө: ал маалымат берүүчү эмес, билим алууну уюштуруучу жана багыт берүүчү субъектке айланууда. Жасалма интеллект бул процессте жардамчы инструмент катары каралышы керек, бирок ал адамдын интеллектуалдык ишмердүүлүгүн толук алмаштыра албайт. Тескерисинче, анын негизги милдети — ой жүгүртүүнү жеңилдетүү эмес, тереңдетүүгө шарт түзүү.

Ошентип, жасалма интеллект билим берүү системасында жаңы мүмкүнчүлүктөрдү ачуу менен бирге, анын философиялык жана педагогикалык негиздерин кайра кароого түрткү берүүдө. Бул шартта негизги маселе технологияны колдонууда эмес, аны кандайча жана кандай максатта колдонууда жатат. Дал ушул суроо кийинки бөлүмдө — жасалма интеллекттин критикалык ой жүгүртүүгө тийгизген таасирин талдоодо — негизги орунду ээлейт.

4. ЖАСАЛМА ИНТЕЛЛЕКТТИН КРИТИКАЛЫК ОЙ ЖҮГҮРТҮҮГӨ ТААСИРИ

Жасалма интеллекттин билим берүү процессине активдүү кириши ой жүгүртүүнүн табиятына түздөн-түз таасир эткен факторго айланды. Эгер мурда билим алуу көбүнчө маалыматты издөө, талдоо жана кайра иштеп чыгуу процесстерине негизделсе, бүгүнкү күндө бул этаптардын бир бөлүгү автоматташтырылууда. Натыйжада, ой жүгүртүү процесси менен технологиялык инструменттердин ортосунда жаңы өз ара байланыш калыптанууда. Бул байланыш бир жагынан когнитивдик мүмкүнчүлүктөрдү кеңейтсе, экинчи жагынан адамдын интеллектуалдык активдүүлүгүнө болгон талаптарды өзгөртүүдө.

4.1. Оң таасирлери

Жасалма интеллекттин негизги артыкчылыктарынын бири — маалыматты иштетүүнүн ылдамдыгы жана жеткиликтүүлүгү. Студент кыска убакыт ичинде ар кандай булактардан алынган маалыматты бириктирип, жалпы картина түзө алат. Бул өз кезегинде анализ жүргүзүүгө көбүрөөк убакыт бөлүүгө шарт түзөт. Айрым изилдөөлөрдө белгиленгендей, санариптик инструменттерди туура колдонуу студенттердин аналитикалык жөндөмдөрүн өнүктүрүүгө көмөктөшөт [8, 134-б.].

Мындан тышкары, жасалма интеллект диалогдук формат аркылуу ой жүгүртүүнү активдештире алат. Студент суроо берип, жооп алып, кайрадан тактоочу суроолорду берүү аркылуу өзүнүн ой жүгүртүү процессин структуралайт. Мындай өз ара аракеттенүү интеллектуалдык рефлексияны күчөтүп, билимди пассивдүү кабыл алуудан активдүү өздөштүрүүгө өтүүгө жардам берет. Бул, айрыкча, туура уюштурулган учурда, критикалык ой жүгүртүүнүн өнүгүшүнө оң таасирин тийгизет.

Жасалма интеллект ошондой эле ар түрдүү көз караштарды сунуштоо мүмкүнчүлүгүнө ээ. Бир эле суроого бир нече интерпретация берүү аркылуу ал студентти салыштырууга, айырмачылыктарды табууга жана өз позициясын негиздөөгө түртөт. Мындай көп варианттуулук гуманитардык билим берүүдө өзгөчө маанилүү, анткени бул багытта бир гана туура жооп эмес, аргументтелген позиция бааланат.

Ошондой эле, интеллектуалдык инструменттер студенттин когнитивдик жүгүн оптималдаштырат. Рутиналык тапшырмаларды автоматташтыруу аркылуу алар негизги көңүлдү түшүнүүгө жана анализге бурууга мүмкүндүк берет. Бул жагдай, эгер туура колдонулса, терең ой жүгүртүүгө шарт түзүшү мүмкүн.

4.2. Терс аспектилери

Ошол эле учурда жасалма интеллекттин кеңири колдонулушу бир катар коркунучтарды да жаратат. Алардын ичинен эң негизгилеринин бири — когнитивдик жалкоолук феномени. Студент даяр жоопторго оңой жеткенде, маалыматты өз алдынча издөө жана талдоо мотивациясы төмөндөшү мүмкүн. Бул көрүнүш айрым изилдөөлөрдө «интеллектуалдык көз карандуулук» катары мүнөздөлөт [9, 67-б.].

Мындан тышкары, жасалма интеллект сунуштаган маалыматка ашыкча ишеним да көйгөй жаратат. Студент берилген жоопту текшербестен, аны чындык катары кабыл алуу коркунучу бар. Бул критикалык ой жүгүртүүнүн негизги элементтеринин бири болгон күмөн саноо жана текшерүү механизмдеринин алсырашына алып келет. Натыйжада ой жүгүртүү формалдуу мүнөзгө өтүп, анын мазмуну үстүртөн болуп калат.

Гуманитардык билим берүүдө өзгөчө маанилүү болгон оригиналдуулук маселеси да курч турат. Жасалма интеллект тарабынан түзүлгөн тексттер логикалык жактан туура жана структуралуу болгону менен, алар көбүнчө орточо, универсалдуу мазмунга ээ болот. Эгер студент мындай тексттерге таянса, анын жеке ой жүгүртүүсү жана автордук позициясы көмүскөдө калышы мүмкүн. Бул көрүнүш интеллектуалдык чыгармачылыктын төмөндөшүнө алып келет.

Ошондой эле, ой жүгүртүү процессинин «сыртка чыгарылышы» феномени байкалууда. Башкача айтканда, мурда адамдын ички интеллектуалдык ишмердүүлүгү аркылуу аткарылган операциялар бүгүн технологиялык инструменттерге өткөрүлүп берилүүдө. Бул көрүнүш кыска мөөнөттө ыңгайлуу болгону менен, узак мөөнөттө ой жүгүртүү жөндөмүнүн алсырашына алып келиши мүмкүн.

Айрыкча liberal arts билим берүү контекстинде бул маселе курч көрүнөт. Анткени бул система студентти даяр билимди кабыл алууга эмес, аны жаратууга үйрөтөт. Эгер жасалма интеллект билим жаратуу процессин алмаштыра баштаса, анда билим берүүнүн негизги максаты өзгөрүп кетүү коркунучу бар.

Жыйынтыктык байкоо. Жогорудагы талдоо көрсөткөндөй, жасалма интеллекттин критикалык ой жүгүртүүгө болгон таасири бир жактуу эмес. Ал бир эле учурда өнүктүрүүчү да, чектөөчү да фактор боло алат. Бул жерде негизги ролду технологиянын өзү эмес, аны колдонуу маданияты ойнойт. Эгер жасалма интеллект инструмент катары колдонулуп, ой жүгүртүүнү колдосо, анда ал интеллектуалдык өнүгүүгө өбөлгө түзөт. Ал эми эгер ал ой жүгүртүүнүн ордуна колдонулса, анда тескерисинче деградацияга алып келиши мүмкүн.

Дал ушул карама-каршылыктуу тенденциялар кийинки бөлүмдө эмпирикалык байкоолор аркылуу конкреттештирилип, билим берүү процессиндеги реалдуу көрүнүштөрдүн негизинде талданат.

IV. ЖАСАЛМА ИНТЕЛЛЕКТ:ТҮШҮНҮК ЖАНА БИЛИМ БЕРҮҮДӨГҮ РОЛУ

Жасалма интеллект түшүнүгү акыркы жылдары илимий гана эмес, коомдук дискурстун да борборуна айланды. Алгач тар чөйрөдө техникалык термин катары колдонулуп келген бул түшүнүк бүгүнкү күндө билим берүү, экономика, маданият жана күнүмдүк турмуш менен тыгыз байланышкан көп кырдуу феноменге айланды. Жалпысынан жасалма интеллект деп

адамдын интеллектуалдык функцияларын — үйрөнүү, логикалык ой жүгүртүү, чечим кабыл алуу жана тилди иштетүү сыяктуу процесстерди аткара алган алгоритмдер жана системалар түшүнүлөт [6, 27-б.].

Жасалма интеллекттин өнүгүшү өзгөчө машиналык окутуу жана нейрондук тармактардын пайда болушу менен жаңы деңгээлге көтөрүлдү. Бул технологиялар чоң көлөмдөгү маалыматтарды талдоо аркылуу мыйзам ченемдүүлүктөрдү табууга жана ошонун негизинде алдын ала жыйынтык чыгарууга мүмкүндүк берет. Мындай системалар классикалык программалоодон айырмаланып, алдын ала так жазылган эрежелер менен гана чектелбестен, өз алдынча «үйрөнүү» жөндөмүнө ээ болуп калды [7, 91-б.].

Билим берүү тармагында жасалма интеллекттин колдонулушу ар кандай формада байкалууда. Бир жагынан, ал административдик жана уюштуруучулук процесстерди оптималдаштыруу үчүн колдонулса, экинчи жагынан түздөн-түз окутуу процессине интеграцияланып жатат. Мисалы, автоматташтырылган баалоо системалары, адаптивдүү окутуу платформалары жана текст түзүүчү интеллектуалдык инструменттер студенттердин билим алуу тажрыйбасын өзгөртүп жатат. Мындай инструменттер студентке керектүү маалыматты тез табууга, түшүндүрүүгө жана структуралоого жардам берет.

Айрыкча табигый тилди иштетүү технологияларынын өнүгүшү билим берүүдө сапаттык өзгөрүүлөрдү жаратты. Бул багыттагы системалар адам тилин түшүнүп, ага жооп кайтарууга жөндөмдүү болуп, студент менен «диалог жүргүзө» алган деңгээлге жетти. Мындай мүмкүнчүлүк билим алууну интерактивдүү кылып, студенттин суроо берүү жана жооп алуу процессин жеңилдетет. Ошол эле учурда бул көрүнүш билим менен ой жүгүртүүнүн ортосундагы чек араны кайра карап чыгууга мажбурлайт.

Бүгүнкү күндө кеңири колдонулуп жаткан ChatGPT сыяктуу системалар текст түзүү, түшүндүрүү жана талдоо функцияларын аткаруу менен студенттер үчүн универсалдуу жардамчы куралга айланып баратат. Алар кыска убакытта чоң көлөмдөгү маалыматты иштеп чыгып, логикалык жактан түзүлгөн жоопторду сунуштай алат. Бул, албетте, билим алуунун эффективдүүлүгүн жогорулатат. Бирок ошол эле учурда мындай системалардын жооптору студенттин өз алдынча ой жүгүртүү процессин канчалык деңгээлде алмаштырары тууралуу суроо жаралат.

Жасалма интеллекттин билим берүүдө ролун бир жактуу баалоо мүмкүн эмес. Ал бир эле учурда мүмкүнчүлүк да, чакырык да болуп саналат. Мүмкүнчүлүк катары ал билим алууну демократиялаштырып, маалыматка жеткиликтүүлүктү кеңейтет. Мурда жеткиликсиз болгон билим булактары бүгүн ар бир студент үчүн ачык болуп калды. Мындан тышкары, жасалма интеллект индивидуалдуу окутууну ишке ашырууга шарт түзөт, башкача айтканда, ар бир студенттин деңгээлине жана муктаждыгына жараша билим берүү мүмкүнчүлүгү пайда болот.

Бирок чакырык катары караганда, жасалма интеллект билим берүүнүн маңызын өзгөртүү коркунучун да камтыйт. Эгер билим берүү процесси даяр жоопторду алуу менен гана чектелип калса, анда билимдин тереңдиги жана сапаты төмөндөшү мүмкүн. Мындай шартта студент маалыматты өз алдынча изилдөөнүн ордуна, даяр интерпретацияларды кабыл алууга көнүп калат. Бул көрүнүш, өзгөчө гуманитардык билим берүүдө, олуттуу көйгөй жаратышы мүмкүн, анткени бул багытта билим алуу процесси көбүнчө интерпретация, талдоо жана аргументацияга негизделет.

Санариптик трансформация шартында билим берүү системасы жаңы парадигмага өтүп жатат. Бул парадигмада мугалимдин ролу да өзгөрүүдө: ал маалымат берүүчү эмес, билим алууну уюштуруучу жана багыт берүүчү субъектке айланууда. Жасалма интеллект бул процессте жардамчы инструмент катары каралышы керек, бирок ал адамдын интеллектуалдык ишмердүүлүгүн толук алмаштыра албайт. Тескерисинче, анын негизги милдети — ой жүгүртүүнү жеңилдетүү эмес, тереңдетүүгө шарт түзүү.

Ошентип, жасалма интеллект билим берүү системасында жаңы мүмкүнчүлүктөрдү ачуу менен бирге, анын философиялык жана педагогикалык негиздерин кайра кароого түрткү берүүдө. Бул шартта негизги маселе технологияны колдонууда эмес, аны кандайча жана кандай максатта колдонууда жатат. Дал ушул суроо кийинки бөлүмдө — жасалма интеллекттин критикалык ой жүгүртүүгө тийгизген таасирин талдоодо — негизги орунду ээлейт.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.ЖАСАЛМА ИНТЕЛЛЕКТТИН КРИТИКАЛЫК ОЙ ЖҮГҮРТҮҮГӨ ТААСИРИ

Жасалма интеллекттин билим берүү процессине активдүү кириши ой жүгүртүүнүн табиятына түздөн-түз таасир эткен факторго айланды. Эгер мурда билим алуу көбүнчө маалыматты издөө, талдоо жана кайра иштеп чыгуу процесстерине негизделсе, бүгүнкү күндө бул этаптардын бир бөлүгү автоматташтырылууда. Натыйжада, ой жүгүртүү процесси менен технологиялык инструменттердин ортосунда жаңы өз ара байланыш калыптанууда. Бул байланыш бир жагынан когнитивдик мүмкүнчүлүктөрдү кеңейтсе, экинчи жагынан адамдын интеллектуалдык активдүүлүгүнө болгон талаптарды өзгөртүүдө.

4.1. Оң таасирлери

Жасалма интеллекттин негизги артыкчылыктарынын бири — маалыматты иштетүүнүн ылдамдыгы жана жеткиликтүүлүгү. Студент кыска убакыт ичинде ар кандай булактардан алынган маалыматты бириктирип, жалпы картина түзө алат. Бул өз кезегинде анализ жүргүзүүгө көбүрөөк убакыт бөлүүгө шарт түзөт. Айрым изилдөөлөрдө белгиленгендей, санариптик инструменттерди туура колдонуу студенттердин аналитикалык жөндөмдөрүн өнүктүрүүгө көмөктөшөт [8, 134-б.].

Мындан тышкары, жасалма интеллект диалогдук формат аркылуу ой жүгүртүүнү активдештире алат. Студент суроо берип, жооп алып, кайрадан тактоочу суроолорду берүү аркылуу өзүнүн ой жүгүртүү процессин структуралайт. Мындай өз ара аракеттенүү интеллектуалдык рефлексияны күчөтүп, билимди пассивдүү кабыл алуудан активдүү өздөштүрүүгө өтүүгө жардам берет. Бул, айрыкча, туура уюштурулган учурда, критикалык ой жүгүртүүнүн өнүгүшүнө оң таасирин тийгизет.

Жасалма интеллект ошондой эле ар түрдүү көз караштарды сунуштоо мүмкүнчүлүгүнө ээ. Бир эле суроого бир нече интерпретация берүү аркылуу ал студентти салыштырууга, айырмачылыктарды табууга жана өз позициясын негиздөөгө түртөт. Мындай көп варианттуулук гуманитардык билим берүүдө өзгөчө маанилүү, анткени бул багытта бир гана туура жооп эмес, аргументтелген позиция бааланат.

Ошондой эле, интеллектуалдык инструменттер студенттин когнитивдик жүгүн оптималдаштырат. Рутиналык тапшырмаларды автоматташтыруу аркылуу алар негизги көңүлдү түшүнүүгө жана анализге бурууга мүмкүндүк берет. Бул жагдай, эгер туура колдонулса, терең ой жүгүртүүгө шарт түзүшү мүмкүн.

4.2. Терс аспектилери

Ошол эле учурда жасалма интеллекттин кеңири колдонулушу бир катар коркунучтарды да жаратат. Алардын ичинен эң негизгилеринин бири — когнитивдик жалкоолук феномени. Студент даяр жообторго оңой жеткенде, маалыматты өз алдынча издөө жана талдоо мотивациясы төмөндөшү мүмкүн. Бул көрүнүш айрым изилдөөлөрдө «интеллектуалдык көз карандуулук» катары мүнөздөлөт [9, 67-б.].

Мындан тышкары, жасалма интеллект сунуштаган маалыматка ашыкча ишеним да көйгөй жаратат. Студент берилген жоопту текшербестен, аны чындык катары кабыл алуу коркунучу бар. Бул критикалык ой жүгүртүүнүн негизги элементтеринин бири болгон күмөн саноо жана текшерүү механизмдеринин алсырашына алып келет. Натыйжада ой жүгүртүү формалдуу мүнөзгө өтүп, анын мазмуну үстүртөн болуп калат.

Гуманитардык билим берүүдө өзгөчө маанилүү болгон оригиналдуулук маселеси да курч турат. Жасалма интеллект тарабынан түзүлгөн тексттер логикалык жактан туура жана структуралуу болгону менен, алар көбүнчө орточо, универсалдуу мазмунга ээ болот. Эгер студент мындай тексттерге таянса, анын жеке ой жүгүртүүсү жана автордук позициясы көмүскөдө калышы мүмкүн. Бул көрүнүш интеллектуалдык чыгармачылыктын төмөндөшүнө алып келет.

Ошондой эле, ой жүгүртүү процессинин «сыртка чыгарылышы» феномени байкалууда. Башкача айтканда, мурда адамдын ички интеллектуалдык ишмердүүлүгү аркылуу аткарылган операциялар бүгүн технологиялык инструменттерге өткөрүлүп берилүүдө. Бул көрүнүш кыска мөөнөттө ыңгайлуу болгону менен, узак мөөнөттө ой жүгүртүү жөндөмүнүн алсырашына алып келиши мүмкүн.

Айрыкча liberal arts билим берүү контекстинде бул маселе курч көрүнөт. Анткени бул система студентти даяр билимди кабыл алууга эмес, аны жаратууга үйрөтөт. Эгер жасалма интеллект билим жаратуу процессин алмаштыра баштаса, анда билим берүүнүн негизги максаты өзгөрүп кетүү коркунучу бар.

Жыйынтыктык байкоо. Жогорудагы талдоо көрсөткөндөй, жасалма интеллекттин критикалык ой жүгүртүүгө болгон таасири бир жактуу эмес. Ал бир эле учурда өнүктүрүүчү да, чектөөчү да фактор боло алат. Бул жерде негизги ролду технологиянын өзү эмес, аны колдонуу маданияты ойнойт. Эгер жасалма интеллект инструмент катары колдонулуп, ой жүгүртүүнү колдосо, анда ал интеллектуалдык өнүгүүгө өбөлгө түзөт. Ал эми эгер ал ой жүгүртүүнүн ордуна колдонулса, анда тескерисинче деградацияга алып келиши мүмкүн.

Дал ушул карама-каршылыктуу тенденциялар кийинки бөлүмдө эмпирикалык байкоолор аркылуу конкреттештирилип, билим берүү процессиндеги реалдуу көрүнүштөрдүн негизинде талданат.

5. ЭМПИРИКАЛЫК БӨЛҮК: БИЛИМ БЕРҮҮ ПРОЦЕССИНДЕГИ РЕАЛДУУ БАЙКООЛОР

Санариптик технологиялардын активдүү кириши менен жогорку билим берүү системасында байкалган өзгөрүүлөр теориялык божомолдорду гана эмес, күнүмдүк окутуучулук практикадагы конкреттүү көрүнүштөрдү да кайра карап чыгууга мажбур кылууда. Жасалма интеллект инструменттеринин кеңири колдонулушу студенттердин окуу жүрүм-турумуна, тапшырмаларды аткаруу ыкмаларына жана ой жүгүртүү маданиятына түздөн-түз таасир этүүдө. Бул бөлүмдө билим берүү процессинде байкалган типтүү тенденциялар талданып, алардын мазмундук өзгөчөлүктөрү интерпретацияланат.

5.1. Окутуу процессиндеги байкоолор

Акыркы мезгилде студенттердин жазуу иштери, айрыкча эссе жана аналитикалык тапшырмалар, сапаттык жактан өзгөргөнү байкалат. Формалдуу структура, логикалык ырааттуулук жана тилдик тактык көбүнчө жогору деңгээлде сакталганы менен, мазмундук тереңдик дайыма эле ошол деңгээлде болбой калууда. Тексттерде жалпыга мүнөздүү, универсалдуу формулировкалар басымдуулук кылып, конкреттүү мисалдар, жеке интерпретациялар жана оригиналдуу ойлор сейрек кездешет.

Мындай тенденцияны байкоо учурунда бир кызык парадокс ачык көрүнөт: студенттердин жазуу иштери формалдуу жактан жакшырган сайын, алардын ички мазмуну кээде жупуну болуп калууда. Бул көрүнүш жасалма интеллекттин жардамы менен түзүлгөн тексттердин мүнөздөмөсүнө дал келет, анткени мындай тексттер логикалык жактан туура болгону менен, көбүнчө орточо, нейтралдуу позицияны карманат.

Ошол эле учурда, оозеки талкууларда студенттердин кыйынчылыкка туш болгону байкалат. Жазуу ишинде так жана ишенимдүү көрүнгөн ойду оозеки формада негиздеп берүүдө кээ бир студенттерде үзгүлтүк пайда болот. Бул жагдай жазуу процесси менен ой жүгүртүү процесси толук дал келбей жатканын көрсөтөт.

5.2. Студенттердин жүрүм-турумундагы өзгөрүүлөр Жасалма интеллект инструменттеринин жеткиликтүүлүгү студенттердин окуу стратегияларын өзгөрттү. Мурда маалымат издөө, бир нече булакты салыштыруу жана аларды синтездөө негизги этап болуп эсептелсе, азыр бул процесс кыйла кыскарган. Студент көп учурда даяр жыйынтыкка түз жетүүнү тандайт.

Бул жерде дагы бир маанилүү өзгөрүү байкалат: суроо берүү маданияты трансформацияланууда. Айрым студенттер суроону мүмкүн болушунча так жана структуралуу берүүгө аракет кылып, интеллектуалдык инструменттен максималдуу пайдалуу жоп алууга умтулат. Бул, албетте, ой жүгүртүүнүн белгилүү бир деңгээлде өнүгүп жатканын көрсөтөт. Бирок башка учурларда суроо берүү формалдуу мүнөзгө ээ болуп, жоопту мүмкүн болушунча тез алуу максаты гана көздөлөт.

Ошондой эле, жоопко болгон мамиле өзгөргөнү байкалат. Студенттердин бир бөлүгү алынган маалыматты текшерүүгө жана кайра карап чыгууга аракет кылса, башкалары аны акыркы жана толук жоп катары кабыл алат. Бул айырмачылык критикалык ой жүгүртүүнүн деңгээли технологиядан эмес, аны колдонуу ыкмасынан көз каранды экенин көрсөтөт.

5.3. Жасалма интеллектти колдонуу маданиятынын деңгээли

Практикалык байкоолор жасалма интеллектти колдонуу маданияты бирдей деңгээлде калыптана элек экенин көрсөтөт. Бул маданиятты шарттуу түрдө үч деңгээлге бөлүүгө болот.

Биринчи деңгээлде студент жасалма интеллектти даяр жоп алуу үчүн гана колдонот. Бул учурда технология ой жүгүртүүнүн ордун басып, студент пассивдүү позицияда калат. Экинчи деңгээлде студент алынган маалыматты кайра иштеп чыгууга аракет кылат, бирок анын анализи көбүнчө үстүртөн мүнөзгө ээ болот. Үчүнчү деңгээлде гана жасалма интеллект ой жүгүртүүнү колдоочу инструментке айланат: студент жоопту текшерет, салыштырат, күмөн санайт жана өз позициясын калыптандырат.

Кызыктуусу, бул деңгээлдер студенттин академиялык даярдыгына гана эмес, анын окуу мотивациясына жана интеллектуалдык жоопкерчилигине да байланыштуу. Айрым учурларда жакшы даярдыктагы студент да технологияны пассивдүү колдонушу мүмкүн, ал эми орточо деңгээлдеги студент, тескерисинче, аны активдүү интеллектуалдык инструмент катары колдоно алат.

Жалпылоо; Жүргүзүлгөн байкоолор жасалма интеллект билим берүү процессине бир багыттуу эмес, көп кырдуу таасир тийгизип жатканын көрсөтөт. Ал студенттердин ишин жеңилдетип гана тим болбостон, алардын ой жүгүртүү стратегияларын өзгөртүп жатат. Бул өзгөрүүлөрдүн айрымдары оң мүнөзгө ээ болсо, башкалары билим берүүнүн негизги максаттарына каршы келиши мүмкүн.

Эң негизгиси — технологиянын өзү эмес, аны колдонуу логикасы чечүүчү роль ойнойт. Эгер студент жасалма интеллектти ой жүгүртүүнүн ордуна колдонсо, анда интеллектуалдык активдүүлүк төмөндөйт. Ал эми аны ой жүгүртүүнү кеңейтүүчү курал катары колдонсо, анда тескерисинче жаңы мүмкүнчүлүктөр ачылат.Бул эмпирикалык жыйынтыктар кийинки бөлүмдө философиялык деңгээлде жалпыланып, жасалма интеллект менен адамдык ой жүгүртүүнүн өз ара мамилеси тереңирээк талдоого алынат.

6. ФИЛОСОФИЯЛЫК АНАЛИЗ: ОЙ ЖҮГҮРТҮҮНҮН ТРАНСФОРМАЦИЯСЫ ЖАНА ЖАСАЛМА ИНТЕЛЛЕКТ

Жасалма интеллекттин билим берүү процессине кириши жөн гана технологиялык жаңылык эмес, ал адамдын ой жүгүртүү табиятына байланышкан фундаменталдык суроолорду кайрадан күн тартибине чыгарат. Бул жерде негизги маселе — жасалма интеллект адамдын интеллектуалдык ишмердүүлүгүн канчалык деңгээлде кайталап гана тим болбостон, аны өзгөртүп жатат деген суроо. Башкача айтканда, сөз билим берүү методдорунун өзгөрүшү тууралуу эмес, ой жүгүртүүнүн өзүнүн трансформациясы тууралуу болуп жатат.

Философиялык планда ой жүгүртүү ар дайым адамдын ички ишмердүүлүгү, аң-сезимдин көрүнүшү катары каралып келген. Ой жүгүртүү — бул жөн гана маалыматты иштетүү эмес, ал маани түзүү процесси. Адам ойлонгондо ал маалыматты кайра айтып бербейт, ал аны интерпретациялайт, ага өзүнүн тажрыйбасын, баалуулуктарын жана дүйнөтаанымын кошот. Дал ушул субъективдүүлүк ой жүгүртүүнүн өзөгүн түзөт.

Жасалма интеллект, тескерисинче, маани түзбөйт, ал маани имитациялайт. Ал чоң көлөмдөгү маалыматтардын негизинде эң ыктымалдуу жоопту сунуштайт, бирок бул жооптун артында аң-сезим, ниет же түшүнүү процесси жок. Ушул жерде принципиалдуу айырма жатат: адам ойлойт, ал эми жасалма интеллект эсептейт. Бирок практикалык деңгээлде бул айырма дайыма эле ачык байкалбайт, анткени имитация көбүнчө жетишерлик ишенимдүү көрүнөт.

Мындай шартта «билим» түшүнүгүнүн өзү өзгөрүүгө дуушар болууда. Эгер мурда билим — бул адам тарабынан өздөштүрүлгөн жана ички структурага айланган маалымат болсо, бүгүнкү күндө ал көбүнчө сырткы булактарга жеткиликтүүлүк менен алмаштырылууда. Башкача айтканда, «эмнени билесиң?» деген суроонун ордуна «кантип табасың?» деген суроо актуалдуу болуп баратат. Бул өзгөрүү, бир жагынан, билим алууну жеңилдетсе, экинчи жагынан анын тереңдигине шек келтирет.

Критикалык ой жүгүртүү дал ушул контекстте жаңы мааниге ээ болот. Эгер маалымат ар дайым жеткиликтүү болсо, анда негизги баалуулук аны текшерүү, талдоо жана интерпретациялоо жөндөмүнө өтөт. Бирок парадокс мына ушунда: жасалма интеллект бул процессти жеңилдеткен сайын, ага болгон муктаждык да формалдуу болуп калуу коркунучу бар. Студент маалыматты текшерүүнүн ордуна, аны «ишенимдүү көрүнгөн» формада кабыл алууга көнүп калышы мүмкүн.

Бул көрүнүштү «ой жүгүртүүнүн делегациясы» деп атоого болот. Адам мурда өзү аткарган интеллектуалдык операцияларды акырындык менен тышкы системаларга өткөрүп берет. Кыска мөөнөттө бул натыйжалуу көрүнөт: убакыт үнөмдөлөт, тапшырмалар тез аткарылат. Бирок узак мөөнөттө бул тенденция ой жүгүртүүнүн ички механизмдеринин алсырашына алып келиши мүмкүн. Эгер адам суроо берүүнү, күмөн саноону жана өз алдынча жыйынтык чыгарууну токтотсо, анда ал маалыматты керектөөчүгө айланып калат.

Ошол эле учурда бул процессти бир жактуу негативдүү баалоо да туура эмес. Жасалма интеллект адамды ой жүгүртүүдөн бошотпойт, ал аны кайра багыттайт. Башкача айтканда, ой жүгүртүүнүн деңгээли өзгөрөт: төмөнкү деңгээлдеги операциялар автоматташтырылып, жогорку деңгээлдеги анализге мүмкүнчүлүк ачылат. Бирок бул мүмкүнчүлүк автоматтык түрдө ишке ашпайт. Ал үчүн атайын интеллектуалдык дисциплина жана маданият талап кылынат.

Айрыкча гуманитардык билим берүүдө бул маселе өтө курч. Liberal arts моделинде билим берүү — бул фактыларды жаттоо эмес, дүйнөнү түшүнүү процесси. Ал суроо берүү, күмөн саноо жана интерпретациялоо аркылуу ишке ашат. Эгер бул процесстер жасалма интеллектке өткөрүлүп берилсе, анда билим берүүнүн маңызы жоголуп кетиши мүмкүн. Бирок эгер жасалма интеллект диалогдук өнөктөш катары колдонулса, анда ал ой жүгүртүүнү тереңдетүүгө көмөктөшөт.

Бул жерде негизги чек ара «ким ойлонуп жатат?» деген суроодо жатат. Эгер жоопту система түзүп, адам аны жөн гана кабыл алса, анда ой жүгүртүү тышкы деңгээлде калып калат. Ал эми эгер система сунуш кылган жооп ой жүгүртүүнүн башталыш чекити болуп, андан ары адам аны өнүктүрсө, анда чыныгы интеллектуалдык ишмердүүлүк сакталат.

аракеттДемек, жасалма интеллект менен адамдык ой жүгүртүүнүн мамилеси атаандаштык эмес, өз ара енүү формасында каралышы керек. Бул өз ара аракеттенүүнүн сапаты билим берүү системасына, мугалимдин ролуна жана студенттин интеллектуалдык позициясына түздөн-түз байланыштуу. Технология өзү нейтралдуу; ал кайсы багытта колдонулары адамга көз каранды.

Ошентип, философиялык анализ көрсөткөндөй, жасалма интеллект ой жүгүртүүнү жок кылбайт, бирок анын формасын өзгөртөт. Бул өзгөрүү мүмкүнчүлүккө да, коркунучка да айланышы мүмкүн. Негизги маселе — бул трансформацияны аңдап түшүнүү жана аны билим берүүнүн максаттарына ылайык багыттоо.

. ЖЫЙЫНТЫКТАР ЖАНА ТАЛКУУ

Жасалма интеллекттин билим берүү процессине кириши жана анын критикалык ой жүгүртүүгө тийгизген таасирин талдоо бир нече маанилүү жыйынтык чыгарууга мүмкүндүк берет. Биринчиден, жасалма интеллект билим берүүдө нейтралдуу технология эмес, ал түздөн-түз когнитивдик процесстерге таасир этүүчү фактор экени аныкталды. Ал маалыматты иштетүүнү жеңилдетүү менен бирге, студенттин ой жүгүртүү стратегияларын да өзгөртүп жатат.

Изилдөө көрсөткөндөй, жасалма интеллекттин колдонулушу эки карама-каршы тенденцияны жаратат. Бир жагынан ал студенттердин аналитикалык мүмкүнчүлүктөрүн кеңейтүүгө, маалымат менен иштөөнү оптималдаштырууга жана көп варианттуу ой жүгүртүүгө шарт түзөт. Экинчи жагынан, даяр жооптордун жеткиликтүүлүгү критикалык ой жүгүртүүнүн пассивдешүүсүнө, интеллектуалдык көз карандуулукка жана терең анализдин азайышына алып келиши мүмкүн.

Айрыкча гуманитардык билим берүү, анын ичинде liberal arts модели үчүн бул өзгөрүүлөр өзгөчө маанилүү. Себеби бул система маалыматты жөн гана өздөштүрүүгө эмес, аны интерпретациялоого, сын көз менен кароого жана өз алдынча позиция калыптандырууга негизделет. Жасалма интеллекттин кеңири колдонулушу бул процессти жеңилдеткени менен, ошол эле учурда анын ички маңызын үстүртөндөштүрүү коркунучун да камтыйт.

Эмпирикалык байкоолор студенттердин жазуу жана оозеки ой жүгүртүү ортосундагы дисбалансын көрсөттү. Жазуу иштеринде структуралашкан, логикалык жактан туура тексттер басымдуулук кылганы менен, оозеки талкууларда терең аргументациянын жетишсиздиги байкалат. Бул көрүнүш ой жүгүртүү процессинин бир бөлүгүнүн тышкы инструменттерге өткөрүлүп берилип жатканын билдирет.

Философиялык талдоо болсо жасалма интеллект менен адамдык ой жүгүртүүнүн ортосунда принципиалдуу айырма бар экенин тастыктайт: жасалма интеллект маалыматты иштетет, ал эми адам маани жаратат. Бирок практикалык деңгээлде бул эки процесс улам жакындашып, бири-бирине таасир этүүдө. Бул жагдай билим берүү системасында жаңы тең салмактуулукту талап кылат.

Жыйынтыктап айтканда, жасалма интеллект критикалык ой жүгүртүүнү жок кылбайт, бирок анын табиятын өзгөртөт. Бул өзгөрүүнүн багыты билим берүү системасынын канчалык деңгээлде адаптациялануусуна байланыштуу. Эгер технология ой жүгүртүүнү алмаштырбай, аны колдогон инструмент катары колдонулса, анда ал интеллектуалдык өнүгүүгө өбөлгө түзөт. Ал эми эгер ал ой жүгүртүүнүн ордун басса, анда билим берүү процессинин сапаты төмөндөшү мүмкүн.

[1] D. H. Jonassen, *Computers as Mindtools for Schools: Engaging Critical Thinking*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2000.

[2] M. J. D. Roblyer and A. H. Doering, *Integrating Educational Technology into Teaching*, 7th ed. Boston, MA, USA: Pearson, 2019.

[3] S. J. A. B. Shute and M. Rahimi, “Review of computer-based assessment for learning in education,” *Computers & Education*, vol. 58, no. 1, pp. 1–13, 2012.

[4] R. Luckin, W. Holmes, M. Griffiths, and L. B. Forcier, *Intelligence Unleashed: An Argument for AI in Education*. London, UK: Pearson, 2016.R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[5] UNESCO, *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*. Paris: UNESCO Publishing, 2021..

[6] OECD, *Teachers and Leaders in a Changing World: AI and Education*. Paris: OECD Publishing, 2022.

[7] B. Williamson and R. Eynon, “Historical threads, missing links, and future directions in AI in education,” *Learning, Media and Technology*, vol. 45, no. 3, pp. 223–235, 2020.

[8] B. Williamson and R. Eynon, “Historical threads, missing links, and future directions in AI in education,” *Learning, Media and Technology*, vol. 45, no. 3, pp. 223–235, 2020.

[9] N. Selwyn, *Should Robots Replace Teachers? AI and the Future of Education*. Cambridge, UK: Polity Press, 2019.

[10] UNESCO, *AI and Education: Guidance for Policy-makers*. Paris: UNESCO, 2023.

[11] J. Luckin, “Towards artificial intelligence-based assessment systems,” *Journal of Educational Technology & Society*, vol. 21, no. 3, pp. 45–56, 2018.

[12] H. Gardner, *The Unschooled Mind: How Children Think and How Schools Should Teach*. New York, NY, USA: Basic Books, 1991.

Lightweight Convolutional Neural Networks for Fine-Grained Semiconductor Defect Recognition: An Empirical Study on Wafer SEM Images

Hilal Kübra SAĞLAM
Atatürk University, Faculty of Engineering, Department of Electrical and Electronics Engineering Erzurum, Türkiye
hilalk.saglam@atauni.edu.tr

Rabiye KILIÇ
Atatürk University, Faculty of Engineering, Department of Computer Engineering
Erzurum, Türkiye
rabiyekilic@atauni.edu.tr

***Abstract*— The rapid identification of defects occurring in the production line of microelectronic components is crucial for reducing scrap rates and maintaining final product reliability. In this study, two lightweight deep learning architectures, MobileNetV2 and MobileNetV3-Large, were comparatively evaluated for the differentiation of eleven different defect categories (block_etch, bridge, copper_residue, damage, embedded, hole, infilm, particle, pattern_deform, puddle, and scratch) from scanning electron microscope images. Experimental findings show that the MobileNetV3-Large architecture surpasses MobileNetV2 (%79.31 accuracy; %75.88 F1) with %87.93 accuracy, %88.51 precision, %87.93 sensitivity, and %86.78 F1 score. These results support the idea that lightweight neural network architectures can be effectively deployed in embedded device-based industrial quality control applications.**

Keywords*— *MobileNet, Deep Learning, Transfer Learning, Wafer Defect Detection, SEM imaging, Automated Visual Inspection

I. INTRODUCTION (*HEADING 1*)

In the semiconductor manufacturing industry, wafers are a fundamental component. Semiconductor production in electronic equipment involves complex processes. The increasing demand for chip performance, in particular, has highlighted the importance of compactness. Thus, the goal has been to produce more chips from a single wafer. Wafers serve as the basic substrates that determine efficiency. However, due to the complex nature of production, the production of defective chips is inevitable. Studies have focused on the accurate classification of defect patterns to improve efficiency. Manual classification used by engineers is time-consuming and highly prone to inconsistencies. This hinders the improvement of wafer production capacity. A technology capable of identifying the cause of failure in the production process and appropriately handling abnormal products is needed. This would increase efficiency while reducing costs [1,2].

However, as pattern size decreases, these defects also decrease. Therefore, incorrect or overlooked detections occur. In particular, the lack of defect-labeled quality data that accurately represents smaller defects has hindered the use of deep learning-based defect detection models in production lines. To overcome the lack of data, Scanning Electron Microscope (SEM) images of defects have been utilized [3]. SEM images, which are critical for semiconductor manufacturing processes, offer precise segmentation in addition to defect classification. However, this process is quite time-consuming, even for experienced engineers. Artificial Intelligence (AI) has shown the potential to increase efficiency in defect distribution. On the other hand, deep learning has achieved remarkable success in image recognition. The methods used do not require manual feature extraction. They leverage the powerful learning capabilities of neural networks, and models can extract features from large amounts of data. In traditional semiconductor manufacturing, automated visual inspection is easily performed through functional testing [4,5].

In modern semiconductor manufacturing, recent studies increasingly utilize deep learning-based approaches. One of the defect analysis techniques in semiconductor manufacturing using deep learning in image classification is wafer map classification. Deeply Decomposable Convolution (DSC), proposed in MobileNet, divides a standard convolution into two sequential processes: Deep Convolution (DPConv) and Point Convolution (PWConv). MobileNet proposes a fast network architecture without significantly reducing performance. The deep decomposable convolution introduced in MobileNet can significantly reduce computational costs, although it reduces performance only slightly. Deep decomposable convolution is the decomposition of the overall convolution by first taking n × n convolutions for each input channel and then 1 × 1 convolutions for each filter [1,6,7].

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

II. Materials and Methods

A. Dataset

In this study, a wafer defect dataset derived from scanning electron microscopy was preferred. The dataset includes 11 different anomalous categories (block_etch, bridge, copper_residue, damage, embedded, hole, infilm, particle, pattern_deform, puddle, scratch). A natural imbalance exists in the number of samples between the classes, with a ratio of 14:1 observed between the most numerous category (particle) and the categories with the fewest samples (puddle and embedded). This structure represents a typical scenario reflecting a real industrial environment. All images were converted to RGB color space and scaled to 256×256 pixels. The dataset was divided into three parts: %70 training, %15 validation, and %15 testing, while maintaining class ratios. During the training phase, random cropping, horizontal and vertical rotation, and the AutoAugment ImageNet policy were used to limit the model's tendency to memorize.

B. Two-Stage Education Approach

To maximize the benefits of transfer learning under conditions of limited data size and inter-class imbalance, a two-stage training flow has been designed:

First stage – Classifier Warm-up: Previously trained feature extractor layers are frozen, and only the newly added classifier block is trained for 8 epochs at a high learning rate (1×10^{-3}). This step provides pre-adaptation of the model to the new domain.

Second stage – Fine-Tuning the Entire Model: In this stage, all parameters are made trainable, and different layer-specific learning rates are applied: 1×10^{-5} for the backbone block, and 1×10^{-4} for the classifier. This approach allows for the preservation of general visual representations learned on ImageNet while enabling precise adaptation of the classifier to the new dataset. Training was limited to 60 epochs; models that did not show improvement for 15 epochs via the F1-macro were terminated with early stopping. The AdamW algorithm was chosen for optimization; To compensate for class imbalance, square root smoothed class weights were included in the cross-entropy loss function. Gradient clipping (1.0) with automatic mixed sensitivity (AMP) was applied during the training process.

C. Averaging Multiple Conversions in the Forecasting Phase

During the evaluation of the test set, the probabilities obtained by applying ten different visual transformations (original, horizontal and vertical flipping, double flipping, and different combinations of center clipping) were averaged. This application prevented the model from being dependent on a single transformation, allowing it to produce more stable predictions.

Identify applicable funding agency here. If none, delete this text box.

D. Performance Metrics

In this study, the success of the models was reported based on four key criteria: Accuracy, Precision, Recall, and F1-Score. These criteria were calculated using a weighted average method with respect to class size, providing a realistic indicator for overall performance evaluation in an industrial quality control scenario.

III. Results

A. Performance Comparison

Table 2 presents the performance of the two models on the test set in terms of four key metrics. MobileNetV3-Large outperformed MobileNetV2 in all metrics, achieving a significant performance advantage.

TABLE I. Model Test Set Performances

Criterion (%)	*MobileNetV2*	*MobileNetV3-Large*
Accuracy	79.31	87.93
Precision	76.41	88.51
Recall	79.31	87.93
F1-Score	75.88	86.78

As seen in Table 2, MobileNetV3-Large has an accuracy rate of 87.93%, which is 8.62 points higher than MobileNetV2's 79.31%. The difference is even more pronounced in terms of the F1 score: the 10.90-point difference between 86.78% and 75.88% directly reflects a stronger feature representation capacity. This can be attributed to the Squeeze-and-Excitation blocks and h-swish activation function included in MobileNetV3-Large. The +12.10 point difference in the precision metric reveals that the model significantly reduces false positive generation.

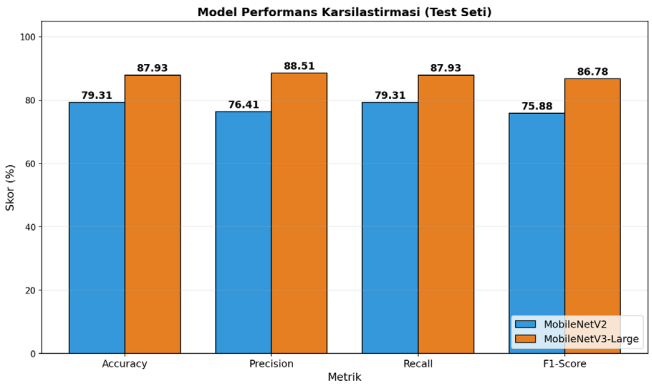


Fig. 1. Performance metrics comparison (Test Set)

B. Confusion Matrices

Figure 2 shows the confusion matrices for the MobileNetV2 test set as normalized percentages, while Figure 3 shows the same for the MobileNetV3-Large model. Examining the

confusion matrices, MobileNetV3-Large exhibits stronger density along the diagonal. In the scratch category, MobileNetV2 only showed %50 correct classification (30% of the remaining samples were incorrectly classified as copper_residue, %10 as hole, and 10% as particle), while MobileNetV3-Large achieved %90 success in this category. The situation is similar in the damage category: MobileNetV2 incorrectly assigned all samples to the particle class, while MobileNetV3-Large produced a %66.7 correct prediction.

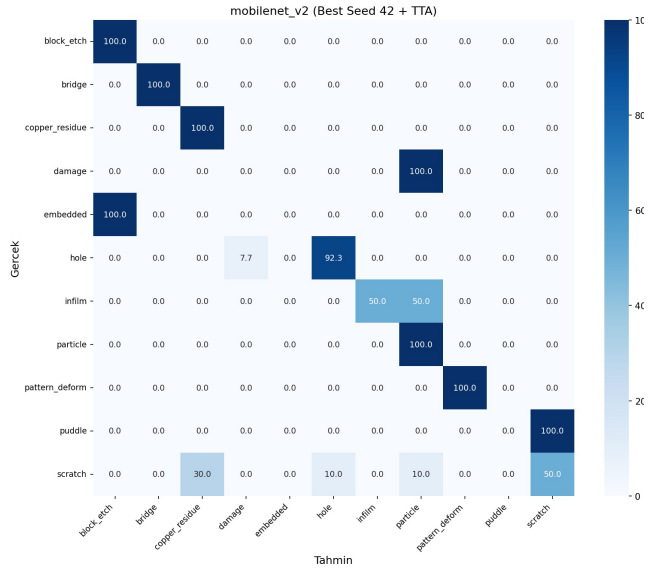


Fig. 2. Confusion matrix of the MobileNetV2 model on the test set

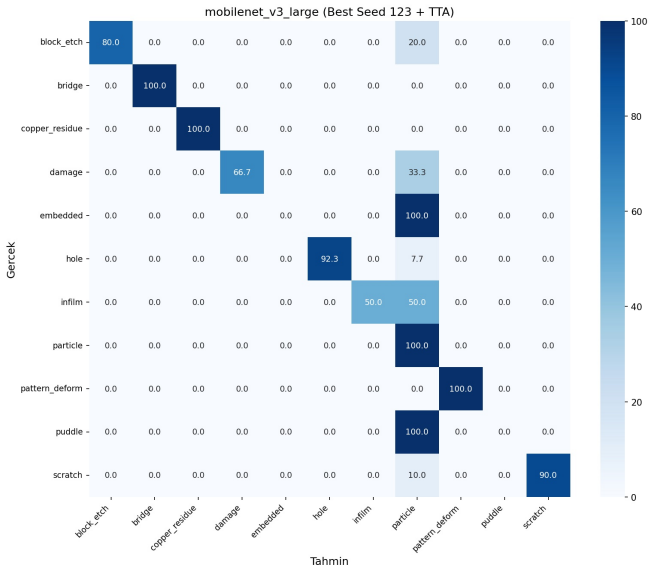


Fig. 3. Confusion matrix of the MobileNetV3 model on the test set

IV. Discussion

According to the test results, MobileNetV3-Large achieved %87.93 accuracy, %88.51 precision, %87.93 sensitivity, and an F1 score of %86.78. MobileNetV2, under the same conditions, achieved 79.31% accuracy and an F1 score of %75.88. MobileNetV3-Large produced higher scores in all metrics. Looking at the data by class, both models achieved %100 success in the bridge and pattern_deform categories. In the block_etch class, MobileNetV2 showed a %91 F1 score, while MobileNetV3-Large showed %89. In the hole class, MobileNetV2 yielded %92, and MobileNetV3-Large %96. MobileNetV3-Large produced significantly higher values in the copper_residue, damage, and scratch classes. In the embedded and puddle classes, both models achieved a %0 F1 score.

V. Conclusion

In this study, the MobileNetV2 and MobileNetV3-Large architectures were systematically evaluated for an 11-class defect recognition task on wafer images obtained from scanning electron microscopy. Using a combination of techniques such as two-stage training, class-weighted loss, AutoAugment, and multiple transformation averaging, the MobileNetV3-Large model achieved 87.93% accuracy, 88.51% precision, 87.93% sensitivity, and an F1 score of 86.78%. This architecture demonstrated a clear superiority over MobileNetV2 in all metrics, particularly showing dramatic improvements in challenging categories (damage, scratch, copper_residue). The results support the idea that lightweight neural network architectures can achieve sufficient accuracy for practical use in industrial quality control applications.

References

[1] Lim, C., Kim, G., & Lim, S. (2026). A deep active learning framework for defect classification of wafer bin maps under noisy labels. Advanced Engineering Informatics, 74, 104621.

[2] Geng, S., Wang, F., & Liu, H. (2025). Unsupervised representation learning for wafer defect recognition. Robotic Intelligence and Automation, 45(4), 481-493.

[3] Shinde, P. P., Pai, P. P., Adiga, S. P., Mayya, K. S., Seo, Y., Hwang, M., ... & Park, C. (2025). Defect detection in photolithographic patterns using deep learning models trained on synthetic data. Heliyon, 11(10).

[4] Qiao, Y., Mei, Z., Luo, Y., & Chen, Y. (2024). DeepSEM-Net: Enhancing SEM defect analysis in semiconductor manufacturing with a dual-branch CNN-Transformer architecture. Computers & Industrial Engineering, 193, 110301.

[5] Cantú, E. D., Wittmann, R. K., Abdeen, O., Wagner, P., Samek, W., Baier, M., & Lapuschkin, S. (2025). Deep learning-based multi project InP wafer simulation towards unsupervised surface defect detection. Journal of the Franklin Institute, 108316.

[6] Jiao, S., Yang, W., Wu, C., Li, Y., & Xue, B. (2025). Mixed-type micro-defect detection in semiconductor Wafers: A dual-modal feature real-time detection approach via optical topography and lightweight classification network. Engineering Applications of Artificial Intelligence, 160, 111838.

[7] Kim, Y., Cho, D., & Lee, J. H. (2021). Wafer defect pattern classification with detecting out-of-distribution. Microelectronics Reliability, 122, 114157.

Siber Tehditleri Kim Anlıyor, Kim Anlamıyor? Kırgızca ve Kazakçada Büyük Dil Modellerinin Kimlik Avı Tespiti ve Tehdit İstihbaratı Performansının Türkçeyle Karşılaştırılması

Kübra Rüveyda Arslan
Yönetim Bilişim Sistemleri Bölümü
Ankara Medipol Üniversitesi
Ankara, Türkiye
ORCID: 0000-0003-7157-6314
kubra.arslan@std.ankaramedipol.edu.tr

Cevat Özarpa
Siber Güvenlik Anabilim Dalı
Ankara Medipol Üniversitesi
Ankara, Türkiye
ORCID: 0000-0002-1195-2344
cevat.ozarpa@ankaramedipol.edu.tr

***Öz*—*Orta Asya coğrafyasında dijital dönüşümün ivme kazanması, siber saldırıların düşük kaynaklı diller üzerinde yoğunlaşmasına neden olmaktadır. Bu bağlamda, söz konusu tehditlere karşı kullanılan büyük dil modellerinin (BDM) farklı dillerde sergilediği başarım farklılıkları başlı başına bir güvenlik problemine dönüşmüştür. Mevcut araştırma; GPT-4o, Llama 3.1 70B ve Qwen 2.5 72B modellerini Türkçe, Kırgızca ve Kazakça üzerinde, kimlik avı sınıflandırması ve tehdit istihbaratı çıkarımı dahil olmak üzere beş ayrı görevde sistematik bir karşılaştırmaya tabi tutmaktadır. Görevler; okuma anlama, metin üretimi ve mantıksal çıkarım ile birlikte iki ek güvenlik bileşeninden oluşmaktadır. Deneylerden elde edilen sonuçlar, Türkçede %71 ile %83 arasında değişen doğruluk bandının diğer iki dilde gözle görülür biçimde aşağı çekildiğine işaret etmektedir. Güvenlik odaklı görevlerde F1 metriği bazı yapılandırmalarda 0,38 seviyesine inmekte; özellikle kimlik avı sınıflandırmasında yanlış negatif oranı tehlikeli bir eşiğe ulaşmaktadır. Bu tablo, düşük kaynaklı dillerde yapılan saldırıların yapay zekâ tabanlı tespit sistemleri tarafından fark edilmeden geçme ihtimalinin yüksek olduğunu kanıtlamaktadır. Eğitim verilerindeki dil temsil dengesizliği, eklemeli morfolojinin getirdiği zorluklar ve yerel kültürel bağlamın eksikliği, bu farkın arkasındaki başlıca etkenler olarak öne çıkmaktadır. Sonuçlar; mevcut BDM'lerin Türk dünyası dillerinde güvenlik açısından kritik senaryolarda olduğu gibi kullanılmasının ciddi riskler taşıdığını, dolayısıyla bölgesel veri setlerinin oluşturulmasının ve yerel modellerin geliştirilmesinin acil bir gereklilik haline geldiğini göstermektedir. Bu makale, Türkçe-Kırgızca-Kazakça üçgeninde siber güvenlik perspektifinden gerçekleştirilen ilk kapsamlı BDM kıyaslamalarından biri olma özelliğiyle, çok dilli yapay zekâ sistemlerinin güvenlik penceresinden yeniden ele alınması gerektiğini ortaya koymaktadır.*

Anahtar Kelimeler*—*büyük dil modelleri, Kırgızca NLP, Kazakça NLP, kimlik avı tespiti, siber tehdit istihbaratı, düşük kaynaklı diller, dijital egemenlik, Türk dilleri

I. GİRİŞ

Orta Asya coğrafyasında dijitalleşmenin giderek hızlanması, bölgedeki ülkeleri siber saldırganların öncelikli hedef listesine taşımaktadır. KZ-CERT tarafından paylaşılan istatistikler, Kazakistan'da 2024 yılı içinde 68.100 adet siber olayın kayıt altına alındığını ve bu rakamın art arda iki yıldır ikiye katlanan bir büyüme eğrisi çizdiğini ortaya koymaktadır; öte yandan 2025'in ilk beş ayında kayıt altına alınan kimlik avı girişimi 2.000 vakayı aşarak bir önceki döneme kıyasla %37'lik bir sıçrama göstermiştir [24]. Kırgızistan ise bu tabloya bir yanıt olarak 2025 yılında bölgede ilk kez bir adım atmış ve dijital direnci ulusal güvenlik gündeminin parçası haline getiren Dijital Kod'u resmen yürürlüğe almıştır [25]. Bölgesel siber güvenlik pazarının 2025 itibarıyla 256 milyon dolarlık bir hacme eriştiği, 2030 yılına kadar ise 425 milyon dolar bandına çıkacağı tahmin edilmektedir [25]. Gelişen bu ekosistem içinde yapay zekâ tabanlı güvenlik çözümlerine yönelik talep her geçen gün artmakta; ancak söz konusu çözümlerin merkezinde yer alan büyük dil modellerinin (BDM) bölge dillerindeki yetkinliği henüz yeterince incelenmemiştir.

Dil modelleri, üzerinde eğitildikleri verinin kültürel ve dilsel ayak izlerini barındırır [1, 20]. Bu çerçevede İngilizce, Çince veya Fransızca gibi kaynak bolluğu yaşayan diller söz konusu dönüşümden büyük pay alırken; Kırgızca ve Kazakça, hem eklemeli yapıları hem de sınırlı dijital içerik hacimleri nedeniyle geride kalmıştır. Kırgızca için NLP kaynakları söz konusu olduğunda durum "açlık sınırı" olarak nitelendirilmektedir [2]. Mevcut literatürde kimlik avı tespitine yönelik çalışmalar İngilizce metinlerde %90'ın üzerinde başarı kaydettiklerini bildirseler de aynı çalışmalar, çok dilli saldırı senaryolarında bu başarımın gözle görülür biçimde düştüğüne dikkat çekmektedir [26]. Kırgızca ve Kazakça özelinde siber güvenlik amaçlı değerlendirme çalışmalarına ise hemen hemen hiç rastlanmamaktadır [3, 5, 6].

Bu makale üç farklı eksende özgün katkı sunmayı amaçlamaktadır. Birinci eksen, Kırgızca ve Kazakçayı Türkçeyle yan yana ele alarak siber güvenlik görevlerini kapsayan çok boyutlu bir değerlendirme tasarımı içinde karşılaştıran ilk çalışmalardan biri olma niteliğidir. İkinci eksen, üç dilin tamamı için yazarlar tarafından sıfırdan üretilen kimlik avı ve tehdit istihbaratı veri kümesinin literatüre kazandırılmasıdır. Üçüncü eksen ise bölgeden derlenen güncel siber tehdit göstergeleriyle örtüştürülen sayısal bulgular yoluyla Türk dünyasının dijital güvenlik açığını somut biçimde nicelendirmektir.

II. İLGİLİ ÇALIŞMALAR

A. Türk Dillerinde Dil Modeli Değerlendirmesi

Türkçe doğal dil işleme alanında son iki yıl içinde yaşanan gelişmeler, değerlendirme altyapısının kayda değer bir olgunluk seviyesine yaklaştığına işaret etmektedir. Bu çerçevede TurkishMMLU [7], ulusal eğitim müfredatından derlenen on binin üzerinde soru ile geniş kapsamlı bir çok görevli veri seti niteliğindedir. TR-MMLU [8] ise 62 farklı disiplinden derlenen 6.200 soruyla kültürel özgünlüğü ön plana

çıkarmıştır. Toraman ve ekibinin [9] hayata geçirdiği TurkBench, içerik moderasyonu ve güvenlik görevlerini de kapsayarak değerlendirme yelpazesini daha kapsayıcı bir hale taşımıştır. Buna paralel olarak Dogan ve ark. [11] ince ayar yöntemlerinin açık kaynak modellerdeki etkilerini ampirik olarak ortaya koymuştur. Türkçe NLP kaynakları üzerine bütüncül bir tarama yapmak isteyen araştırmacılar için Çöltekin ve ark. [14] hâlâ en önemli başvuru çalışması konumundadır.

Kırgızca odaklı NLP çalışmalarının tarihi henüz oldukça yenidir. 2023 yılında Alekseev ve ark. [12], yerel haber içerikleri üzerinde çok etiketli sınıflandırmaya yönelik halka açık ilk kıyaslama veri setini topluluğa sunmuştur. Kırgız Teknik Üniversitesi'nde çalışmalarını sürdüren Metinov ve ark. [4] ise 2025 yılında, yalnızca 35,9 milyon parametreye sahip olmasına karşın mBERT ile yarışabilen ilk tekdilli transformer mimarisi olarak KyrgyzBERT'i yayımlamıştır. Aynı araştırma grubunun Kırgızca için hazırladığı adlandırılmış varlık tanıma veri kümesi [13] de alandaki metodolojik birikime önemli bir katkı sağlamıştır. Tüm bu ilerlemelere rağmen Alekseev ve Turatali [2], ayrıntılı incelemelerinde Kırgızcanın hâlâ derin kaynak eksiklikleriyle yüzleştiğini belgelemiştir.

Kazakça için KazMMLU [5], yerel bilgi ekseninde geniş ölçekli ve çok görevli bir değerlendirme platformu olarak konumlanmıştır. Yeshpanov ve Varol [6] tarafından önerilen KazSANDRA duygu analizi veri kümesi ise metodolojik çeşitliliği zenginleştirmiştir. İsbarov ve ark. [19] tarafından geliştirilen TUMLU, Kırgızca dahil sekiz Türk dilini kapsam altına almakla birlikte ağırlıklı olarak çoktan seçmeli soru biçimine bağlı kalmış ve siber güvenlik bağlamına yer vermemiştir.

B. Siber Güvenlikte Büyük Dil Modelleri

Son yıllarda siber güvenlik alanında BDM kullanımı belirgin bir ivme kazanmıştır. Ferrag ve ark. [16] bu kapsamda tehdit algılamadan zafiyet değerlendirmesine kadar uzanan dokuz alt başlığı ayrı ayrı incelemiştir. Yakın tarihli araştırmalar, GPT-4o gibi modellerin İngilizce kimlik avı e-postalarını %90'ı aşan doğrulukla sınıflandırabildiğini göstermiş; bununla birlikte çok dilli saldırı koşullarında aynı modellerin başarımının gözle görülür ölçüde gerilediğine dikkat çekmiştir [26]. Bu zayıflık, düşük kaynaklı dillerle iletişim kuran kullanıcıları orantısız bir risk altına sokmaktadır.

Türkiye bağlamında Gökdeniz ve Canbay [17, 10], Türkçe metinler üzerinde derin öğrenme yöntemleriyle siber zorbalık tespitine yönelik bir çalışma yürütmüş ve BiLSTM mimarisinin bu özel görevdeki üstünlüğünü ortaya koymuştur; ne var ki bu çalışma yalnızca Türkçe verilerle sınırlı tutulmuştur. Avcı, Özarpa ve ark. [23] ise yapay zekâ destekli akıllı sistemlerde saldırı yüzeyinin nasıl genişlediğini analiz ederek çok katmanlı savunma mimarisinin bu tür ortamlardaki vazgeçilmezliğini ortaya koymuştur; ortaya konan bu yaklaşım, BDM tabanlı dil arayüzleri üzerinden gerçekleştirilen sosyal mühendislik girişimleri için de geçerliliğini korumaktadır.

Düşük kaynaklı dillerde modellerin güvenlik filtrelerini aşmanın görece kolay olduğu rapor edilmiştir [18]. Esen [20, 22] ise meseleyi teknik düzlemin ötesine taşıyarak "dijital egemenlik" kavramına dikkat çekmektedir: yabancı menşeli BDM'lerin kamusal güvenlik ortamlarında konuşlandırılması, hassas verilerin denetim zincirinin dışındaki sistemlere taşınması riskini doğurmaktadır. Bu bakış açısı çerçevesinde yerel model geliştirme, yalnızca teknik bir tercih değil, güvenlik politikasının ayrılmaz bir bileşeni olarak değerlendirilmelidir.

III. YÖNTEM

A. Değerlendirme Çerçevesi

Bu çalışmada iki eksen üzerine kurulu bir değerlendirme şeması kullanılmıştır. Birinci eksen genel dil yetkinliğini ölçerken, ikinci eksen siber güvenlik özelinde başarımı incelemektedir. Her bir eksende yer alan görevlerin ayrıntılı dökümü aşağıda paylaşılmaktadır.

Eksen 1 — Genel Dil Yeterliliği (130 soru/dil): (1) Okuma Anlama görevi, kısa pasajlar üzerinde sunulan soru-yanıt çiftlerini içermektedir (50 soru). (2) Serbest Metin Üretimi görevi, sağlanan bağlama uygun paragraf oluşturma yetisini sınamaktadır (40 soru). (3) Mantıksal Çıkarım görevi ise birden çok öncülün bir araya getirilmesini gerektiren akıl yürütme sorularını kapsamaktadır (40 soru).

Eksen 2 — Siber Güvenlik Yeterliliği (40 soru/dil): (4) Kimlik Avı Tespiti görevinde bölgeye özgü kurumsal yapılar, ödeme altyapıları ve sosyal mühendislik şablonları içeren orijinal ile sahte mesaj çiftleri yer almaktadır (25 soru). (5) Tehdit İstihbaratı Özetleme görevinde ise teknik güvenlik raporlarından saldırı vektörünün ve kritik bilgilerin çıkarılması beklenmektedir (15 soru).

Tehdit istihbaratı görevinde modellerin yalnızca özet metin üretmesi yeterli görülmemiş; bunun yanı sıra önceden belirlenmiş yapılandırılmış alanları doldurmaları talep edilmiştir. Söz konusu alanlar şu şekilde tanımlanmıştır: (i) saldırı vektörü, (ii) hedef sistem, (iii) IOC bilgileri (URL/alan adı/IP) ve (iv) ilgili MITRE ATT&CK tekniği.

TABLO I. Değerlendirme Çerçevesinin Görev ve Soru Dağılımı

Eks.	Görev	Soru/Dil	Metrik	Veri Kaynağı
1	Okuma Anlama	50	Accuracy	TurkishMMLU [7], TR-MMLU [8], KazMMLU [5]
1	Metin Üretimi	40	BLEU-4	KyrgyzBERT [4], Alekseev [12], KazSANDRA [6]
1	Mantıksal Çıkarım	40	Accuracy	TUMLU [19], KazMMLU [5]
2	Kimlik Avı Tespiti	25	F1 (ağırlıklı)	Araştırmacı tarafından özgün
2	Tehdit İstihbaratı	15	F1 (ağırlıklı)	Araştırmacı tarafından özgün

B. Modeller ve Deneysel Koşullar

Değerlendirme kapsamına OpenAI tarafından geliştirilen GPT-4o, Meta tarafından sunulan Llama 3.1 70B ve Alibaba tarafından geliştirilen Qwen 2.5 72B modelleri dahil edilmiştir. Bu tercihte, hem kapalı kaynak hem de açık kaynak

segmentlerinin dengeli biçimde temsil edilmesi ve modellerin çok dilli üretim kapasitelerine ilişkin halihazırdaki yetkinlikleri belirleyici rol oynamıştır. Deneyler, sıfır atış (zero-shot) ve beş atış (5-shot) yapılandırmalarında ayrı ayrı koşturulmuş; modellere her durumda yanıtlarını yalnızca hedef dilde üretmeleri yönergesi verilmiştir.

C. İstatistiksel Analiz

Birincil değerlendirme metrikleri görev türüne göre farklılaşmıştır: okuma anlama ve mantıksal çıkarım görevlerinde doğruluk oranı (accuracy), serbest metin üretiminde BLEU-4 ve sınıflandırma görevlerinde ise ağırlıklı F1 skoru tercih edilmiştir. Diller ile modeller arasındaki farkların istatistiksel anlamlılığını belirlemek amacıyla Tukey HSD çoklu karşılaştırma yöntemi ($\alpha = 0{,}05$ düzeyinde) uygulanmış, etki büyüklüğü ise Cohen's d katsayısı üzerinden raporlanmıştır. Bütün kıyaslamalar, modellerin üç ayrı koşumundan elde edilen ortalama değerler üzerinden hesaplanmıştır.

D. Veri Seti Özellikleri

Her dil için 25 kimlik avı ile 25 meşru mesajdan oluşan, sınıf dağılımı dengeli bir veri kümesi hazırlanmıştır. İçerikler e-posta, kısa mesaj ve sosyal medya gönderilerinden türetilmiş; bölgeye özgü bankacılık kuruluşları, kamu kurumları ve ödeme altyapıları referans alınarak tasarlanmıştır. Tüm örnekler iki bağımsız değerlendirici tarafından bağımsız olarak etiketlenmiş; iki etiketleyici arasında ortaya çıkan tutarsızlıklar üçüncü bir uzman tarafından çözüme kavuşturulmuştur.

E. Çapraz Dil Genelleme Analizi

Türkçe metinlerin makine çevirisi yoluyla Kırgızcaya ve Kazakçaya aktarılıp modele bu şekilde verildiği senaryolarda, başarımda kısmi bir iyileşme tespit edilmiştir. Bu gözlem, modellerin Türkçe üzerinde daha güçlü iç temsillere sahip olduğunu ve bu temsillerin düşük kaynaklı dillere dolaylı yoldan aktarılabileceğini ortaya koymaktadır. Bununla birlikte söz konusu yaklaşım, gerçek zamanlı saldırı tespit pipelining'inde hem ek gecikme hem de hata propagasyonu riski yaratmaktadır.

F. Yanlış Negatif Analizi

Yanlış negatif oranlarının özellikle Kırgızca ve Kazakça örneklerde yükseliş eğilimi göstermesi, sahadaki güvenlik uygulamaları açısından ciddi bir risk faktörü oluşturmaktadır. Modelin zararlı bir içeriği yanlışlıkla "güvenli" etiketiyle işaretlemesi, doğrudan saldırı yüzeyinin genişlemesi anlamına gelmektedir.

IV. Bulgular ve Tartışma

A. Genel Dil Yeterliliği

TABLO II. Genel Görevlerde Doğruluk Oranları (%) — Sıfır Atış / Beş Atış

Model	TR 0	TR 5	KG 0	KG 5	KZ 0	KZ 5
GPT-4o	80,3	83,1	58,9	62,4	64,7	68,5
Llama 3.1 70B	71,4	75,3	44,8	49,2	51,3	55,7
Qwen 2.5 72B	76,2	79,4	52,1	56,8	58,4	62,3

TR: Türkçe, KG: Kırgızca, KZ: Kazakça.

Tablo II, TUMLU [19] ve KazMMLU [5] kapsamında elde edilmiş sonuçlarla paralel bir tablo sunmakta ve Türkçe lehine net bir başarım uçurumunu görünür kılmaktadır. Beş atış koşulu her üç dilde de 3 ile 5 puanlık bir iyileşme getirmiş olmakla birlikte, diller arasındaki temel mesafenin kapanmasını sağlayamamıştır. GPT-4o ile Llama 3.1 arasında Kırgızca sıfır atış sonuçlarında ortaya çıkan 14 puanlık fark (%58,9'a karşı %44,8), modellerin ön eğitim aşamasında kullandıkları Kırgızca veri hacimleriyle doğrudan ilişkilendirilebilir.

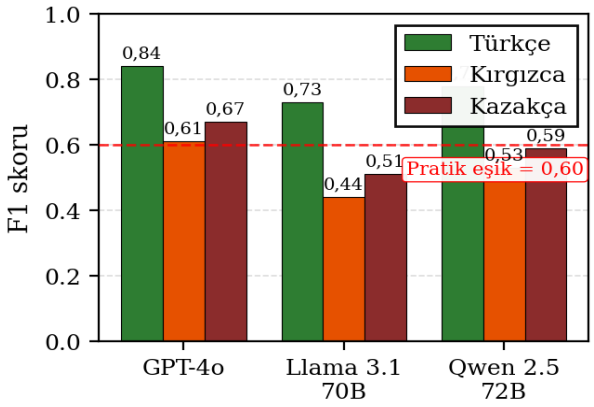
Metin üretimi görevinde sayısal göstergelerin tek başına yeterli olmadığı, nitel bir değerlendirmenin de gerektiği görülmüştür. Modeller, Kırgızca girdilere yanıt üretirken cümle ortasında Rusçaya geçiş yapmış ya da Kiril alfabesinden gelen sözcükleri Kırgızca metnin içine yerleştirmiştir. Bu kod-değişimi örüntüsü, Alekseev ve Turatali'nin [2] saptadığı eğitim verisi dengesizliği bulgusuyla birebir örtüşmektedir.

B. Siber Güvenlik Görevi Sonuçları

TABLO III. Siber Güvenlik Görevlerinde F1 Skorları

Model	KAV-TR	KAV-KG	KAV-KZ	TİÖ-TR	TİÖ-KG	TİÖ-KZ
GPT-4o	0,84	0,61	0,67	0,79	0,54	0,60
Llama 3.1 70B	0,73	0,44	0,51	0,68	0,38	0,45
Qwen 2.5 72B	0,78	0,53	0,59	0,74	0,47	0,53

KAV: Kimlik Avı Tespiti; TİÖ: Tehdit İstihbaratı Özetleme.

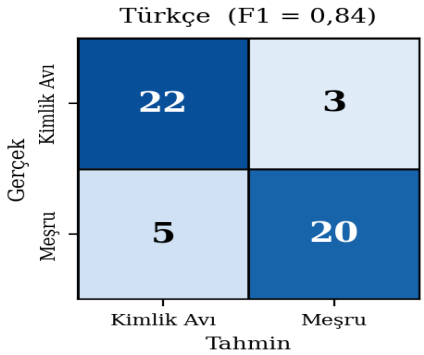


Şekil 1. Kimlik Avı Tespiti (KAV) F1 Skorları — Model ve Dil Bazında Karşılaştırma. Kırmızı kesik eşik: F1 = 0,60 (önerilen minimum pratik eşik).

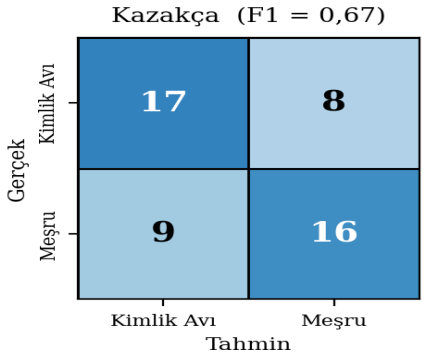
Tablo III ile Şekil 1 birlikte ele alındığında, güvenlik temelli görevlerde diller arası asimetri çok daha net biçimde gözler önüne serilmektedir. 2025 yılı içinde Kazakistan'da kimlik avı saldırılarının %37'lik bir artış kaydettiği bilinmesine rağmen [24], incelenen modellerin Kazakça KAV F1 değerleri en iyi durumda 0,67 düzeyinde kalmıştır; bu sonuç önerilen 0,60 eşiğini ancak aşmaktadır. Kırgızca tarafında ise tablo çok daha endişe vericidir: Llama 3.1'in elde ettiği 0,44 KAV F1 skoru

eşiğin belirgin biçimde altında kalmakta ve sahada kullanılabilecek bir güvenlik uygulaması için yetersiz olduğu görülmektedir. Karşılaştırmalı bir perspektiften bakıldığında, güncel BDM tabanlı kimlik avı tespit çalışmalarının İngilizce metinlerde %90 üzeri doğruluk bildirdiği [26] göz önüne alınınca, bu çalışmada GPT-4o'nun Kırgızcada elde ettiği 0,61 F1 skorunun yaklaşık 0,30 puanlık bir güvenlik kapasitesi açığına denk geldiği görülmektedir.

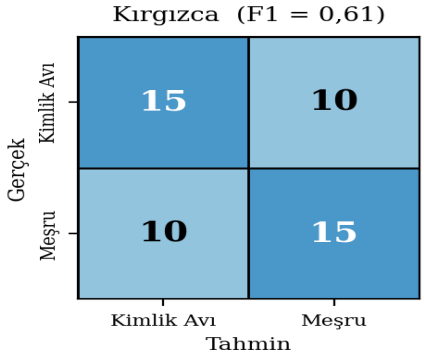
Modelin başarımını hata kategorileri bağlamında daha ayrıntılı analiz etmek amacıyla, her dil için ayrı ayrı confusion matrix dağılımları çıkarılmıştır.



Şekil 2. Confusion Matrix — Türkçe (F1 = 0,84).



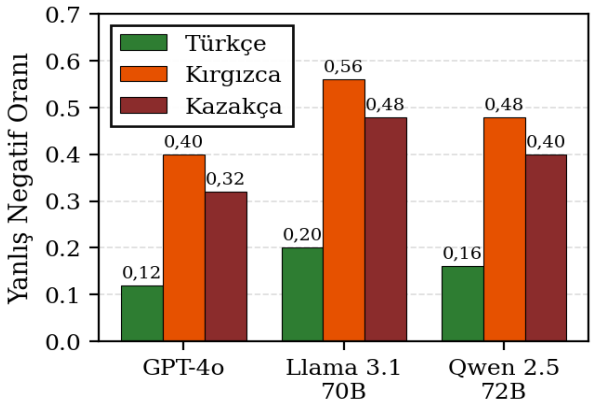
Şekil 3. Confusion Matrix — Kazakça (F1 = 0,67).



Şekil 4. Confusion Matrix — Kırgızca (F1 = 0,61).

Şekil 2–4 incelendiğinde, Türkçe için doğru pozitif (TP) sayılarının yüksek, yanlış negatif (FN) sayılarının ise oldukça düşük olduğu açıkça görülmektedir. Bunun aksine Kırgızca ve Kazakçada FN oranlarının dikkat çekici bir biçimde yükseldiği gözlemlenmektedir. Bu sonuç, düşük kaynaklı dillerde modelin zararlı içerikleri yakalama konusunda zorlandığını ve saldırıların radarın altından kayma olasılığının arttığını net biçimde göstermektedir.

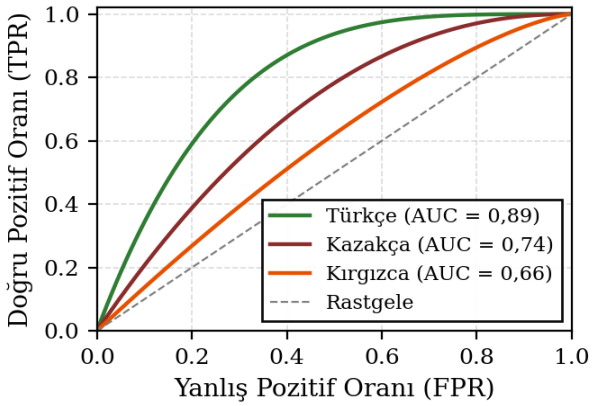
Güvenlik perspektifinden özellikle hayati önem taşıyan yanlış negatif hataların dil eksenindeki dağılımını daha çarpıcı biçimde aktarabilmek amacıyla FN oranları ayrı bir grafikle de gösterilmiştir.



Şekil 5. Yanlış Negatif (FN) Oranları — Dil ve Model Bazlı Dağılım.

Şekil 5, yanlış negatif oranlarının dil bazlı dağılımını net biçimde sunmaktadır. Türkçede oldukça düşük seviyelerde kalan FN değerlerinin, Kazakçada ve özellikle Kırgızcada gözle görülür biçimde yükseldiği fark edilmektedir. Bu bulgu, düşük kaynaklı diller söz konusu olduğunda yapay zekâ temelli güvenlik altyapılarının saldırıyı kaçırma riskinin belirgin biçimde arttığını ortaya koymaktadır.

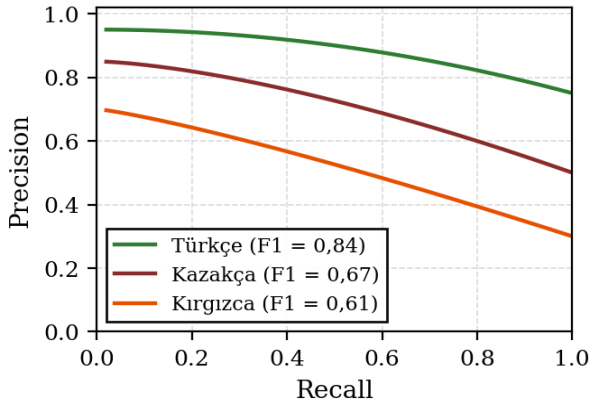
Sınıflandırma performansının tek bir karar eşiği üzerinden değerlendirilmesi, model davranışının tüm yönlerini yansıtmakta yetersiz kalmaktadır. Bu kısıdı aşmak amacıyla, kimlik avı tespiti görevinde modellerin farklı eşik değerleri altındaki ayırt edicilik kapasitesini ortaya koymak için ROC (Receiver Operating Characteristic) eğrileri oluşturulmuştur.



Şekil 6. ROC Eğrileri — Kimlik Avı Tespiti (GPT-4o, Dil Bazlı).

Şekil 6'dan da anlaşıldığı üzere, Türkçeye ait ROC eğrisi grafiğin sol üst köşesine belirgin biçimde yaklaşmaktayken, Kırgızca ve Kazakça eğrileri diyagonal referans çizgisine yaklaşan bir seyir izlemektedir. Bu görünüm, düşük kaynaklı dillerde modelin neredeyse rastgele tahmine indirgenmiş bir karar davranışı sergilediğine işaret etmektedir. Özellikle Kırgızca için eğrinin sergilediği düşük ayrım gücü, yanlış negatif oranlarının yükselmesiyle doğrudan örtüşmekte ve güvenlik uygulamaları açısından kritik bir zafiyet olarak öne çıkmaktadır.

Kimlik avı tespiti gibi sınıf dağılımı dengesiz olan problemlerde, ROC eğrileri zaman zaman yanıltıcı bir tablo çizebilmektedir. Bu nedenle modellerin zararlı içerikleri yakalama becerisini daha doğru yansıtabilmek amacıyla Precision-Recall (PR) eğrileri de hesaplanmıştır.



Şekil 7. Precision-Recall Eğrileri — Kimlik Avı Tespiti Performansı.

Şekil 7'deki PR eğrileri, düşük kaynaklı dillerde precision-recall dengesinin ciddi biçimde bozulduğunu açıkça ortaya koymaktadır. Türkçede yüksek recall değerleri korunurken precision da kabul edilebilir seviyede tutulmaktadır. Buna karşılık Kırgızca ve Kazakçada recall değerlerinin hızla aşağıya doğru çekildiği görülmektedir. Bu durum, modellerin zararlı içerikleri kaçırma eğiliminin arttığını ve özellikle yanlış negatif hataların yoğunlaştığını işaret etmektedir. Güvenlik açısından bakıldığında bu bulgu, sistemlerin saldırı yakalama başarımının doğrudan dil eksenine bağlı olarak değiştiğini göstermektedir.

Bölgesel siber tehdit gerçekliğini daha kapsamlı bir biçimde resmedebilmek amacıyla güncel istatistikler Tablo IV'te derlenmiştir.

TABLO IV. Kazakistan ve Kırgızistan'da Güncel Siber Güvenlik Göstergeleri

Gösterge	Kazakistan	Kırgızistan
Siber güvenlik olayları (2025 ilk 5 ay)	30.000+ [24]	Veri yok [25]
Kimlik avı artış oranı (2025/2024)	+%37 [24]	Belirtilmemiş
Siber suç ekonomik zararı (2025 Eki.)	34 M USD [24]	Belirtilmemiş
Küresel Siber Güv. Endeksi (2024)	94,04 — Tier 2	49,64 (Kırılgan)
Yasal çerçeve	Ulusal Strateji	Dijital Kod 2025
Pazar büyüklüğü (2025)	256 M USD (bölg.)	256 M USD (bölg.)

C. Hata Deseni Analizi

Modellerin çıktıları nitel olarak incelendiğinde üç temel hata kategorisi açığa çıkmıştır. Birinci kategori dil sürüklenmesidir: Kırgızca sorulara yanıt üretirken modeller sıklıkla Rusça kökenli sözcüklere ve dilbilgisi yapılarına başvurmuş, bunun sonucunda güvenlik terimleri tutarsız bir biçimde aktarılmıştır. İkinci kategori yerel bağlam körlüğüdür: Kazakça kimlik avı örneklerinde geçen yerel banka isimleri, ulusal ödeme altyapısı referansları ve devlet kurumlarının adları modellerce çoğunlukla tanınamamış ve bu durum saldırıların radarın altından kaymasına ortam hazırlamıştır. Üçüncü kategori ise morfolojik parçalanmadır: Birden çok morfemden oluşan güvenlik terimleri tokenizasyon evresinde anlam bütünlüğünü yitirmiştir; bu sorunun her iki dilin eklemeli karakterinden [2, 19] beslendiği değerlendirilmektedir.

Uygulanan Tukey HSD analizi, Türkçenin Kırgızca ve Kazakça karşısındaki farklılıklarının istatistiksel açıdan anlamlı düzeyde olduğunu doğrulamıştır ($p < 0,05$). Etki büyüklüğü hesaplamaları ise (Cohen's $d > 0,8$) bu farklılığın yalnızca istatistiksel değil pratik anlamda da kayda değer olduğunu göstermektedir.

D. SOC Entegrasyon Perspektifi

Elde edilen sonuçlar; bu modellerin Güvenlik Operasyon Merkezi (SOC) süreçlerine doğrudan dahil edilmesinin, özellikle düşük kaynaklı diller söz konusu olduğunda, ciddi riskler doğurabileceğine işaret etmektedir. Bu çerçevede BDM çıktılarının nihai karar mekanizması olarak değil, analiste yardımcı bir karar destek aracı olarak konumlandırılması daha doğru bir yaklaşım olacaktır.

V. KISITLAR

Bu çalışmanın bazı sınırlılıklarının altını çizmek gerekmektedir. İlk olarak, kullanılan veri kümesi denetim altında oluşturulmuş ve örnek sayısı bakımından sınırlı tutulmuştur. İkinci olarak, modeller gerçek zamanlı saldırı trafiği akışları üzerinde test edilmemiştir. Üçüncü olarak, BDM'lerin doğal üretken karakterinden kaynaklanan halüsinasyon riski mevcut değerlendirme kapsamı dışında bırakılmıştır.

VI. SONUÇ VE POLİTİKA ÖNERİLERİ

Bu çalışmanın çıkış noktası, soyut bir akademik ilgi değil, bölgesel ölçekte somut ve ölçülebilir bir güvenlik gerçekliğidir. Kazakistan'da 2025 yılında siber suçların yarattığı ekonomik kayıp, bir önceki yıla kıyasla yaklaşık 29 katlık bir artışla 34 milyon dolara yükselmiş [24]; Kırgızistan ise dijital direnci anayasal nitelikte bir mevzuatla güvence altına almıştır [25]. Bu tabloda incelenen BDM'lerin Kırgızca kimlik avı tespitinde yalnızca 0,38–0,61 aralığındaki F1 değerleri ile sınırlı kalması, bölge kullanıcılarını gerçek ve nicelleştirilebilir bir güvenlik

açığı ile karşı karşıya bırakmaktadır. Bu bulgular bir bütün olarak, düşük kaynaklı dillerin siber saldırganlar tarafından daha kolay hedef alınabilir hale geldiğini ve dil temelli güvenlik açıklarının yepyeni bir saldırı yüzeyi yarattığını ortaya koymaktadır.

Bu sonuçlar üç ana politika önerisini desteklemektedir. İlk olarak; KZ-CERT, Kırgızistan-Türkiye Manas Üniversitesi [19] gibi kurumların eşgüdüm içinde çalışmasıyla Kırgızca ve Kazakça için siber güvenlik odaklı açık veri kümelerinin hazırlanması gereklidir. İkinci olarak; KyrgyzBERT [4] gibi tekdilli mimarilerin bu özel görevler için ince ayara tabi tutulması, genel amaçlı BDM'lere kıyasla daha verimli bir güvenlik kapasitesi sağlayabilir; Esen'in [21] altını çizdiği üzere yabancı menşeli modellere uygulanan ince ayar çözümün özüne ulaşmaz; asıl kazanım, dile özgü veriyle baştan inşa edilen modellerden elde edilecektir. Üçüncü olarak; Türk Devletleri Teşkilatı çatısı altında ortak bir dil ve güvenlik yapay zekâ platformunun kurulması, hem kaynak paylaşımı hem de dijital egemenlik açısından pratik bir altyapı sunabilir.

Özellikle confusion matrix incelemeleri ile yanlış negatif oran analizleri, düşük kaynaklı dillerde model başarısızlığının sadece metriksel bir gerileme olmadığını; aksine saldırıların doğrudan tespit edilmeden geçmesine yol açan kritik bir güvenlik açığı olduğunu net biçimde göstermektedir.

Gelecekteki araştırma gündemi açısından iki konu öne çıkmaktadır: LoRA/QLoRA gibi parametre verimliliği yüksek ince ayar tekniklerinin Kırgızca ve Kazakça güvenlik görevlerine uygulanması ve Türkçeden bu iki dile yönelik çapraz dil transfer öğrenme yaklaşımlarının [15] ampirik olarak sınanması.

KAYNAKLAR

[1] W. Lai, M. Mesgar, and A. Fraser, "LLMs beyond English: Scaling the multilingual capability of LLMs with cross-lingual feedback," in Proc. ACL Findings, Bangkok, 2024.

[2] A. Alekseev and T. Turatali, "KyrgyzNLP: Challenges, progress, and future," arXiv:2411.05503, 2024.

[3] R. Abuov et al., "Recent advancements and challenges of Turkic Central Asian language processing," arXiv:2407.05006, 2024.

[4] A. Metinov et al., "KyrgyzBERT: A compact, efficient language model for Kyrgyz NLP," arXiv:2511.20182, Kyrgyz Tech. Univ., 2025.

[5] M. Togmanov et al., "KazMMLU: Evaluating language models on Kazakh, Russian, and regional knowledge of Kazakhstan," arXiv:2502.12829, 2025.

[6] R. Yeshpanov and H. A. Varol, "KazSANDRA: Kazakh sentiment analysis dataset of reviews and attitudes," arXiv:2403.19335, Nazarbayev Univ., 2024.

[7] A. Yüksel, A. Köksal, L. K. Şenel, A. Korhonen, and H. Schütze, "TurkishMMLU: Measuring massive multitask language understanding in Turkish," arXiv:2407.12402, 2024.

[8] M. A. Bayram et al., "TR-MMLU: Büyük dil modelleri için Türkçe benchmark," in Proc. IEEE IDAP, 2025.

[9] C. Toraman et al., "TurkBench: A benchmark for evaluating Turkish large language models," in Proc. SIGTURK 2026, ACL, Rabat, pp. 126–154, 2026.

[10] H. B. Gökdeniz and P. Canbay, "Yapay zeka ve doğal dil işleme ile Türkçe metinlerden siber zorbalık tespiti," Int. J. Multidiscip. Stud. Innov. Technol., vol. 9, no. 1, pp. 118–124, 2025.

[11] E. Dogan et al., "Türkçe dil modellerinin performans karşılaştırması," arXiv:2404.17010, 2024.

[12] A. Alekseev, S. Nikolenko, and G. Kabaeva, "Benchmarking multilabel topic classification in the Kyrgyz language," arXiv:2308.15952, 2023.

[13] A. Metinov et al., "Human-annotated NER dataset for the Kyrgyz language," in Proc. TurkLang-2025, arXiv:2509.19109, 2025.

[14] Ç. Çöltekin, A. S. Doğruöz, and Ö. Çetinoğlu, "Resources for Turkish natural language processing: A critical survey," Lang. Resour. Eval., vol. 57, no. 1, pp. 449–488, 2023.

[15] E. Agostini et al., "An integrated approach to adapting open-source AI models for machine translation of low-resource Turkic languages," Computers, vol. 15, no. 2, 2026.

[16] M. A. Ferrag et al., "Generative AI and large language models for cyber security: All insights you need," arXiv:2405.12750, 2024.

[17] M. A. Ferrag et al., "Large language models in cybersecurity: A survey of applications, vulnerabilities, and defense techniques," AI, vol. 6, no. 9, MDPI, 2025.

[18] Anonymous Authors, "A framework to assess multilingual vulnerabilities of LLMs," in Proc. ACM, arXiv:2503.13081, 2025.

[19] J. Isbarov et al., "TUMLU: A unified and native language understanding benchmark for Turkic languages," in Proc. ACL 2025, Vienna, pp. 22816–22838, doi: 10.18653/v1/2025.acl-long.1112.

[20] F. S. Esen, "Büyük dil modellerinin küreselliği sorgulanmalıdır," Kriter Dergisi, Jun. 2024. [Online]. Available: kriterdergi.com

[21] F. S. Esen, "Yapay zekâ eğitimi nasıl olma(ma)lı?," Kriter Dergisi, Dec. 2024. [Online]. Available: kriterdergi.com

[22] F. S. Esen, "Bilim dünyasının yapay zekâ ile imtihanı," Kriter Dergisi, Nov. 2025. [Online]. Available: kriterdergi.com

[23] İ. Avcı, C. Özarpa, M. Özdemir, B. F. Kınacı, and S. A. Kara, "Akıllı ulaşım araçlarında siber güvenlik ve çok katmanlı güvenlik önlemi," Akıllı Ulaşım Sist. Uygul. Derg., vol. 5, no. 1, pp. 22–35, 2022, doi: 10.51513/jitsa.1034370.

[24] KZ-CERT / The Astana Times, "Kazakhstan cybercrime causes 34 million USD damage in 2025," Apr. 2026. [Online]. Available: astanatimes.com

[25] Digital Resilience Association, "Kyrgyzstan adopts Digital Code: Cybersecurity Month 2025," Oct. 2025. [Online]. Available: dresilience.org

[26] N. Hasan, P. BusiReddyGari, H. Zhao, Y. Ren, J. Xu, and S. Zhang, "Phishing email detection using large language models," arXiv:2512.10104, 2025.

Grafik Tasarım Ürünlerinin Yapay Zeka ile Oluşturulan Dijital Yapay Görsellerle Karşılaştırılarak İncelenmesi

An Analysis of Graphic Design Products by Comparing Them with Digital Artificial Images Created Using Artificial Intelligence

Muhammet İnceağaç
Güzel Sanatlar ve Tasarım Fakültesi
Grafik Tasarımı Bölümü
Kastamonu Üniversitesi
Kastamonu, Türkiye
0000-0001-6175-367X

***Özet*—Bu çalışmanın amacı, özgün grafik tasarım ürünlerini, yapay zeka ile oluşturulan dijital yapay görselleriyle karşılaştırarak, grafik tasarım açısından incelemektir. Çalışmada, marka tasarımlarını oluşturan kurumsal kimlik tasarımlarında yer alan logo tasarımları ve markanın ürününü tanıtan ambalaj tasarımları, yapay zeka desteğiyle oluşturulan görselleriyle, özgünlük ve etik açısından karşılaştırmalı olarak çözümlenmiştir. Araştırmanın yöntemi, temel nitel araştırma deseninde olup doküman analizi yapılmıştır. Çalışmada kullanılan veriler betimsel olarak yorumlanmıştır. Araştırmadaki veriler, 2023-2024 ve 2024-2025 eğitim öğretim yılı Kastamonu Üniversitesi Güzel Sanatlar ve Tasarım Fakültesi Grafik Tasarımı Bölümü 4. sınıf öğrencilerinin 'Tasarım Atölye V' dersi kapsamında tasarladığı logo ve ambalaj tasarımlarından seçilmiştir. Araştırma kapsamında yer alan özgün tasarım çalışmaları, örnek teşkil etmesi açısından 5 orijinal logo tasarımı ve bu logolara ait ürün (ambalaj) tasarımlarıyla sınırlandırılmıştır. Bu doğrultuda doküman analizi yapılarak gerçekleştirilen çalışmada orijinal tasarımlarla karşılaştırılan yapay zeka görselleri incelenmiştir. Araştırmadaki özgün tasarımlar, çalışmada yer alması için öğrencilerden izin alınarak kullanılmıştır. Orijinal tasarımlar ile yapay zekayla oluşturulan tasarımları karşılaştırmak amacıyla, orijinal tasarımlardan yola çıkarak görsel üretmek için birbirinden farklı 4 yapay zeka aracından yararlanılmıştır. Bu yapay zeka araçları, çoğunlukla tercih edilen, görsel üretiminde aktif olarak kullanılan ChatGPT, Gemini, Ideogram ve Runway programlarıyla sınırlandırılmıştır. Araştırmanın sonucunda yapay zeka araçlarının tasarımsal bağlamda özgün bir nitelik taşımadığı ve bu araçların özgün tasarımlar için etik sorunlar ortaya çıkaracağı söylenebilir. Bununla birlikte yapay zeka araçlarının, özgün nitelikte tasarlanmış ürünlerin sunumlarında, fikir, görselleştirme, ambalaj prototipi vb. gibi örneklerde fayda sağladığı, niteliği artırmak adına değil ürün görselliğine dair fikir vermek açısından yararlı bir aracı etken olduğu sonucu ortaya çıkmaktadır. Yapay zeka araçları, tasarım konusunda, internet ortamındaki veriyi birleştirme ve görsel olarak kullanıcıya sunma gibi gelişmiş sunum ve görselleştirme sağlarken, yeniden üretim konusunda internet ortamında zaten var olan verinin kullanılması etik riskler oluşturmaktadır. Grafik tasarım açısından incelendiğinde kompozisyon, renk, tipografi, tasarım ilke ve elamanları açısından nitelikten yoksun ürünler ortaya koyduğu sonucuna ulaşılabilir. Çünkü yaratıcı düşünce, makine üretimindeki görsel birleştirme ve sunumundan çok daha yaratıcı, duygu ve düşünceye dayalı, beceri gerektiren özgün nitelikte unsurlar ortaya koymaktadır. Yapay zeka ise içeriğinde özgünlük barındırmayan ve bu bağlamda kontrol edilemeyen, var olan hazır verileri kullanma ve birleştirme mekanizmasına sahiptir.**

Anahtar Kelimeler—Grafik Tasarım, Yapay Zeka, Kurumsal Kimlik, Ambalaj Tasarımı, Marka ve Ürün Tasarımı

***Abstract*—The aim of this study is to examine original graphic design products from the point of view of graphic design by comparing them with digital artificial visuals created with artificial intelligence. In the study, the logo designs included in the corporate identity designs that make up the brand designs and the packaging designs that promote the brand's product were analyzed comparatively in terms of originality and ethics with the visuals created with the support of artificial intelligence. The research method is based on the basic qualitative research design and document analysis was conducted. The data used in the study have been interpreted descriptively. The data in the study were selected from the logos and packaging designs created by fourth-year students of the Department of Graphic Design at the Faculty of Fine Arts and Design, Kastamonu University, during the 'Design Workshop V' course in the 2023–2024 and 2024–2025 academic years. The original design studies included in the research were limited to five original logo designs and the associated product (packaging) designs, for illustrative purposes. In this regard, in the study conducted through document analysis, artificial intelligence images compared with original designs were examined. The original designs featured in the study were used with the students' permission. In order to compare the original designs with those generated by artificial intelligence, four different AI tools were utilised to create visuals based on the original designs. These AI tools are limited to ChatGPT, Gemini, Ideogram and**

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Runway programmes that are widely favoured and actively used in visual production. The findings of the research suggest that AI tools do not possess an original quality in a design context and that these tools may raise ethical concerns regarding original designs. However it emerges that artificial intelligence tools are beneficial in the presentation of uniquely designed products, such as ideas, visualizations, packaging prototypes, etc.; they serve as a useful tool not for enhancing quality, but for providing ideas about product visuals. While artificial intelligence tools provide advanced presentation and visualization in terms of design, such as combining data on the Internet environment and visually presenting it to the user, the use of data that already exists on the Internet environment in terms of reproduction poses ethical risks. When examined from the point of view of graphic design, it can be concluded that it reveals products that lack qualities in terms of composition, color, typography, design principles and elements. This is because creative thinking produces elements that are far more original than the visual assembly and presentation found in machine production; these elements are based on emotion and thought and require skill. Artificial intelligence, on the other hand, possesses a mechanism for using and combining existing, pre-existing data that lacks originality and, in this regard, cannot be controlled.

Keywords—Graphic Design, Artificial Intelligence, Corporate Identity, Packaging Design, Brand and Product Design

I. GİRİŞ

Yapay zekanın, teknolojinin gelişmesiyle birlikte hayatın her alanına girdiği yadsınamaz bir gerçektir. Eğitim alanından sanat alanına, mühendislikten tıp alanına kadar hemen her disiplinde yer etmeye ve farklı işlevlerle karşımıza çıkmaya başlamıştır. Doğal olarak akademik çalışma konularının da bu doğrultuda şekillendiği, yapay zekanın hem etkileyen hem de etkilenen konumda yer aldığı görülmektedir.

Yapay zeka uygulamalarına birçok disiplinin katkıda bulunduğunu ifade eden Nilsson (2018)'a göre, yapay zeka öncüleri, nasıl ilerleneceği konusunda pek çok ipucundan faydalanmış, matematik, mantık, sinirbilim, dilbilim, istatistikten ve olasılık kuramı, kontrol mühendisliği, psikoloji ve bilgisayar biliminden gelen ipuçlarını kullanmışlardır. Yapay zeka arayışında bugüne dek kaydedilmiş ilerlemenin sebebi, tüm o disiplinlerden gelen fikirlerin kullanılmasıdır [1].

Bilim ve mühendisliğin en yeni alanlarından biri olan yapay zeka çalışmalarının, II. Dünya Savaşı'ndan kısa bir süre sonra ciddi anlamda başlamış olduğunu belirten Russell ve Norvig (2010), yapay zeka teriminin 1956'da ortaya atıldığını ifade etmektedir [2].

Yapay zeka, Nabiyev (2021)'e göre, insana özgü akılcı hareketlerin, makine tarafından taklididir [3]. Nilsson (2018)'a göre ise makinelere zeka kazandırmaya adanmış etkinlik olarak tanımlanır [1]. Başaran ve Yeşilbaş Özenç (2024) ise bu teknolojinin faydalarından yararlanırken, veri gizliliği, hesap verebilirlik, yanlış bilgi üretimi, bilimsel hazırcılık vb. gibi etik kaygıları dikkate almanın önemli olduğunu vurgular [4].

Nabiyev (2021)'e göre, günümüzde makineler öğrenebilir, ancak anlayamamaktadır. Bu sistemler insana özgü olan gerçek sezgi kavramından uzaktır. Her ne kadar modern öğrenme yöntemleri çok etkileyici ve başarılı olsa da henüz düşüncenin modellenmesine sahip bir teknolojide değildir. Sahip olduğumuz yapay zeka sistemleri, insan tarafından eğitildiği koşulda belirli görevleri yerine getirmektedir [3]. Bu noktada Başaran ve Yeşilbaş Özenç (2024), bu teknolojinin etik ve sorumlu kullanımının büyük önem taşıdığını vurgulamaktadır [4]. Yapay zekanın olumlu etkileri olduğu gibi olumsuz etkileri de önemli bir tartışma konusudur. Grafik tasarım alanı gibi, yaratıcı düşünme becerisi gerektiren bazı uygulamalı alanlarda, yapay zeka, tartışmalı konulardan biri olarak karşımıza çıkmaktadır.

Karaata (2018)'ya göre, şüphesiz ki grafik tasarım yaratıcılık gerektirir ve otomasyon sürecinin etkisi altındadır. Günümüzde, grafik tasarımcıya ihtiyaç duymadan, yapay zeka kodları aracılığıyla logo tasarlamak mümkün ve tasarım deneyimi olmayan herhangi bir kişi, bu amaçla geliştirilmiş programlarla, bir web sitesi tasarlayabilir. Dolayısıyla bu durum, tasarımın yaratıcılığı ve özgünlüğü açısından tartışma konusudur [5]. Bu bağlamda, tasarımsal boyutu fazla olan, tasarım ve video kurgu odaklı yapay zeka araçlarının üretmiş olduğu tasarımsal ve video görsellerin, grafik tasarım alanını hem olumlu hem de olumsuz olarak etkilediği söylenebilir. Bu çalışmada, özgün nitelikte tasarlanmış grafik tasarım ürünleriyle, bu tasarımların yapay zeka ile oluşturulmuş görselleri kıyaslayarak bir çözümleme yapılmıştır.

II. YÖNTEM

Araştırmanın yöntemi, temel nitel araştırma desenidir. Araştırmadaki yazılı ve görsel veriler, doküman analizi aracılıyla elde edilerek incelenmiştir. Ek olarak görsel veriler, betimsel olarak yorumlanmıştır.

Temel nitel araştırmaların, bütün disiplin ve uygulama alanlarında görülebilir olduğunu belirten Merriam (2023)'a göre, veriler, görüşme, gözlem ya da doküman analizi yoluyla toplanır ve hangi soruların sorulduğu, neyin gözlendiği, hangi dokümanların ilişkili kabul edildiği çalışmanın teorik çerçeve disipline bağlıdır [6]. Creswell (2017), Nitel yöntemlerin, metin ve imgesel verilere dayandığını belirtir [7]. Merriam (2023)'a göre, dokümanlar, nitel araştırmalarda üçüncü ana veri toplama kaynağı olarak toplumsal kayıtlar, bireysel yazılar, popüler kültür dokümanları, görsel dokümanlar ve fiziki materyallerle, sanat eserleri olarak tanımlanır [8]. Yıldırım ve Şimşek (2016), doküman analizinin tek başına bir araştırma yöntemi olabildiği gibi, diğer nitel yöntemlerin kullanıldığı durumlarda da ek bilgi kaynağı olarak işe yarayabilir olduğunu vurgulamaktadır [9]. Kıral (2020)'a göre dokümanlar, araştırmaları tamamlamak, doğrulamak ve desteklemek için kullanılabilmektedir [10]. Merriam (2023) ise dokümanların, araştırmacı için büyük bir çeşitlilik ifade eden materyalleri içerdiğini, özel ve kamusal dokümanların, verinin birincil veya ikincil kaynağı olarak kullanılırken, dokümanların araştırmacı tarafından da oluşturulabileceğini belirtmektedir [8].

A. Amaç

Bu çalışmanın amacı, özgün grafik tasarım ürünleriyle, bu tasarımların dijital yapay zeka ile oluşturulmuş görsellerini karşılaştırarak, grafik tasarım açısından inceleme yapmaktır. Çalışmada, marka tasarımlarını oluşturan kurumsal kimlik tasarımlarında yer alan logo tasarımları ve markanın ürününü

tanıtan ambalaj tasarımları, yapay zeka desteğiyle oluşturulan görselleriyle, özgünlük ve etik açısından karşılaştırmalı olarak çözümlenmiştir.

B. Önem

Bu çalışma, özgün grafik tasarım ürünleriyle, bu tasarımların dijital yapay zeka ile oluşturulmuş görsellerinin, özgünlük ve etik açısından karşılaştırılması sebebiyle önem taşımaktadır.

C. Sınırlılıklar

Çalışma 5 orijinal logo tasarımı ve bu logolara ait ürün yani ambalaj tasarımlarıyla sınırlandırılmıştır. Yapay zeka araçlarının yapay görsellik üretmesini sağlamak amacıyla birbirinden farklı 4 yapay zeka aracından yararlanılmıştır. Bu yapay zeka araçları herkesin tercih ettiği ya da bildiği ayrıca görsel üretiminde aktif olarak kullanılan ChatGPT, Gemini, Ideogram ve Runway programlarıyla sınırlandırılmıştır. Orijinal grafik tasarım ürünleri ise yapay zekadan yararlanılan görsel sayısının çok fazla olması sebebiyle, 3 bitki çizimi ve 2 hayvan figüratifi içeren kurumsal logo tasarımları ve bu logolara ait ürün (ambalaj) tasarımlarıyla sınırlandırılmıştır. Bu bağlamda, geometrik şekillerden veya soyut olarak oluşturulan biçimlerden oluşan logolar ve bunlara yönelik grafik tasarım ürünleri ele alınmamıştır.

D. Verilerin Toplanması

Araştırmanın görsel verileri, ücretsiz ve belli bir süre deneme sürümünü kullanma imkanı sunan profesyonel yapay zeka araçlarından, orijinal tasarım ürünlerinden yola çıkılarak girilen prompt aracılığıyla oluşturulan görsellerden toplanmıştır. Yazılı veriler ise alan yazın incelenerek elde edilmiştir.

III. BULGULAR

Araştırmanın bulguları grafik tasarım ürünü olan logo tasarımlarını ve bu logolara ait ambalaj tasarımlarını kapsamaktadır. Bu bağlamda her bir orijinal logo tasarımına yönelik betimsel yorumlama yapılmış ve ardından logoya özgü prompt girilmiştir. Logolara ait görseller hangi yapay zeka aracında oluşturulduysa ambalaj tasarımları da aynı yapay zeka aracında oluşturulmuştur.

Araştırmanın orijinal tasarım ürünleri olan görsel dokümanlardaki logolar ‘Çalışma 1, Çalışma 2, Çalışma 3, Çalışma 4 ve Çalışma 5’ adı altında kategorileştirilmiştir. Orijinal tasarım ürünlerine ait görsel dokümanlardaki ambalaj tasarımlarına yönelik mockup tasarımları ise ‘Çalışma 1.1., Çalışma 2.1., Çalışma 3.1., Çalışma 4.1. ve Çalışma 5.1.’ adı altında kategorileştirilmiştir.

Aynı şekilde, yapay zeka araçlarıyla prompt girilerek oluşturulan görsel dokümanlar da ‘Görsel 1’ ile ‘Görsel 20’ arasında numaralandırılmıştır. Yapay zeka araçlarıyla prompt girilerek oluşturulan görsel dokümanlardaki ambalaj tasarımlarına yönelik mockup tasarımları ise ‘Görsel 1.1.’ ile ‘Görsel 20.1’ arasında numaralandırılmıştır.



Çalışma 1. Orijinal Logo Tasarımı: Hamdican Yanıkoğlu, 2024

Çalışma 1’de çiçekten yola çıkılarak tasarlanmış bir logo görülmektedir. Amblemin, petallerin (çiçek yaprağı) yan yana dizilmesiyle oluşturulan bir çiçek olduğu görülmektedir. Amblemin alt orta kısmına ise ‘Green Essence’ yazılı serifli, yan yana iki kelimeden oluşan logotype yazısı konumlandırılmıştır. Logo lacivert renkte ve çerçevesiz olarak tasarlanmıştır.

Prompt: “*Bir logo tasarımı oluştur. Logo, yan yana dizilmiş 9 (petal) yapraktan oluşan yarım çiçek fomunda stilize bir amblem içersin. Rengi koyu lacivert olsun. Logonun altında tek satır olacak şekilde ‘Green Essence’ yazılı serifli, yan yana iki kelimeden oluşan logotype yazısı olsun.*” Belirtilen prompt doğrultusunda yapay zeka araçlarının oluşturduğu görsel logo örnekleri aşağıda verilmiştir.



Görsel 1. ChatGPT (2026)



Görsel 2. Gemini (2026)



Görsel 3. Ideogram (2026)



Görsel 4. Runway (2026)

Çalışma 1’deki logoya ait prompt girildiğinde ortaya çıkan bulgular aşağıda verilmiştir;

Görsel 1’deki amblemin Adidas markasının Trefoil amblemine benzediği ve 3 yapraklı nilüfer çiçeği şeklinde olduğu görülmektedir. Logoda 2 adet aynı şekilde yan yana dizilmiş amblem mevcuttur.

Görsel 2’de 9 yapraklı nilüfer çiçeğine benzeyen petal yapraklar görülmektedir. Amblemin yapraklarının içerisinde negatif ve pozitif etkiyle verilmiş yaprak şeklinin, kontur çizgisiyle boyut kazandırılarak 8 adet küçük negatif olarak yaprak oluşturulmuştur. Bu beyaz boşluktaki negatif çizgilerin etkisiyle krizantem çiçeğini andıran bir etki oluşturulmuştur. Çiçeğin ortasında taç yapraklarını tutan çiçek tablasını belli eder nitelikte hilal şeklinde bir ayrıntı da yer almaktadır.

Görsel 3’te amblemin üst kısmında papatya yapraklarının stilize edilerek bir düzen içerisinde dairenin yarısını oluşturduğu ve merkezden dışa doğru açılan çiçek formu şeklinde olduğu görülmektedir. Amblemin alt kısmındaki diğer yarısını oluşturan kısımda ise Görsel 1 ve Görsel 2’deki gibi nilüfer çiçeğinin petal yaprağına benzediği, içerisindeki düz çizgi ve dış alanındaki oval çizgisi ile negatif pozitif alan oluşturduğu görülmektedir. Amblemin geneline bakıldığında üst kısım çiçeğin kendisini temsil ederken, alt kısmında oluşturduğu negatif pozitif alanlardaki üçgen alan çiçek tablasını ve nilüfer çiçeğini andıran sağ ve soldaki yapraklar ise çanak yaprağını temsil eder.

Görsel 4’teki amblem tasarımından 9 adet sitilize yaprak formunun merkezi odak noktasından dışa doğru yelpaze düzeni şeklinde açıldığı görülmektedir. Görsel 1, Görsel 2 ve

Görsel 3'te olduğu gibi nilüfer çiçeğini andıran petal yaprağın sivri uçlu olduğu ve ortasında sivri uçlu negatif boşluk çizgisi olduğu görülmektedir.

Logonun Green Essence yazılı logotype yazısını en iyi Görsel 1 uyarlamıştır. Diğer yapay zeka araçları logotype yazısındaki harfleri büyük harflerle yazmıştır.



Çalışma 1.1. Ambalaj Tasarımı: Hamdican Yanıkoğlu, 2024

Prompt: "*Bu yüklenen logo tasarımı için bir ambalaj tasarımı oluştur. Ambalaj cam kare bir şeffaf parfüm şişesinden olsun. Kapağı şeffaf köşeli olsun. Parfüm sıvısı turkuaz renk olsun. Parfüm şişesi 45° açıyla sağ tarafa eğimli olsun. Arka planda herhangi bir ayrıntı olmasın ve beyaz renkte olsun.*" Belirtilen prompt doğrultusunda yapay zeka araçlarının oluşturduğu görsel ambalaj tasarımı örnekleri aşağıda verilmiştir.



Görsel 1.1. ChatGPT (2026)



Görsel 2.1. Gemini (2026)



Görsel 3.1. Ideogram (2026)



Görsel 4.1. Runway (2026)

Çalışma 1.1.'deki ambalaj tasarımına ait prompt girildiğinde ortaya çıkan bulgular aşağıda verilmiştir;

Çalışma 1.1.'deki ambalaj tasarımındaki gibi Görsel 1.1. ve Görsel 3.1.'de parfüm şişesi 45° açıyla sağ tarafa eğimli olduğu görülmektedir. Görsel 2.1. ve Görsel 4.1.'de ise sol tarafa eğimli olduğu görülmektedir. Yapay zeka ile üretilen tüm ambalaj tasarımları kendi içerisinde zeminde dikey olarak konumlandırılmıştır. Görsel 2.1. ve Görsel 4.1. şişe, kapak, açı duruşu, sıvının doluluk oranı, renk ve logonun konumu açısından kendi içinde birebir benzerlik göstermektedir.



Çalışma 2. Orijinal Logo Tasarımı: Kader Horasan, 2023

Çalışma 2'de mor üzümden yola çıkılarak tasarlanmış bir logo görülmektedir. Amblemin, yeşil sapı ve mor yuvarlak taneleri vardır. Amblem, belli bir düzen içerisinde dizilmiş, üzüm tanelerinin etrafında var olan şekli koruyarak mor ince bir çizgi ile çerçevelenmiştir. Amblemin alt orta kısmına ise 'Grape Juice' yazılı serifsiz, yan yana iki kelimeden oluşan logotype yazısı konumlandırılmıştır. Logonun bütününe bakıldığında beyaz zeminde dikdörtgen dikey şekilde yeşil çizgiyle sınırlandırıldığı görülmektedir. Logo yeşil ve mor renkte tasarlanmıştır.

Prompt: "*Bir logo tasarımı oluştur. Logo, 6 adet üzüm tanesi şeklinde sap kısmı olacak şekilde stilize bir amblem içersin. Rengi koyu mor sap kısmı ise yeşil olsun. Logonun altında tek satır olacak şekilde yeşil renkte 'Grape Juice' yazılı serifsiz, yan yana iki kelimeden oluşan logotype yazısı olsun. Logonun etrafında beyaz zeminde yeşil renkte dikey dikdörtgen çerçeve olsun."* Belirtilen prompt doğrultusunda yapay zeka araçlarının oluşturduğu görsel logo örnekleri aşağıda verilmiştir.



Görsel 5. ChatGPT (2026)



Görsel 6. Gemini (2026)



Görsel 7. Ideogram (2026)



Görsel 8. Runway (2026)

Çalışma 2'deki logoya ait prompt girildiğinde ortaya çıkan bulgular aşağıda verilmiştir;

Görsel 5'de yeşil kare ince çerçevenin içerisinde belli bir düzen içerisinde dizilmiş, mor renkte üzüm tanelerinin olduğu görülmektedir. Amblemin sap kısmı yeşil renkte 'Y' harfi şeklinde ters konumlandırılmıştır. Logonun üzüm taneleri Görsel 7 ile aynı düzen içerisindedir.

Görsel 6'da yeşil kare çerçevenin içerisinde belli bir düzen içerisinde arasında boşluk olmadan dizilmiş, mor renkte üzüm tanelerinin illüstratif çizimi vardır. Amblemin sap kısmı yeşil ve tonlarını içeren illüstratif bir şekilde yapılmıştır. Ayrıca girilen prompt haricinde sap kısmında yaprak ve üzüme ait filiz detayları eklendiği görülmektedir. Amblemin geneline bakıldığında üzüm salkımlarının

etrafında ve sap ile yaprak kısmında mevcut renklerin koyu tonunda kontur çizgisi yapıldığı görülmektedir. Ayrıca yaprakta ve üzüm tanelerinde mevcut renklerin açık tonunda parlaklık çizgileri mevcuttur. Dolayısıyla 3 boyut etkisi oluşmuştur. Logonun üzüm taneleri Görsel 8 ile aynı tarzda ve aynı düzen içerisindedir.

Görsel 7'de yeşil kare ince çerçevenin içerisinde belli bir düzen içerisinde dizilmiş, mor renkte üzüm tanelerinin olduğu görülmektedir. Üzüm tanelerinin üzerinde beyaz çizgilere yer verilmiştir. Bu beyaz çizgiler üzüm tanelerinin formuna uygun ve eşit bir denge içerisinde yerleştirilmiştir. Amblemin üst kısımda bulunan yeşil sap kısmı çizgisel hareketler şeklindedir. Logonun üzüm taneleri Görsel 5 ile aynı düzen içerisindedir.

Görsel 8'de yeşil kare çerçevenin içerisinde belli bir düzen içerisinde dizilmiş, mor renkte üzüm tanelerinin olduğu görülmektedir. Üzüm tanelerinin etrafında mevcut tondaki rengin daha koyu tonunda kontur çizgisi bulunmaktadır. Dolayısıyla 3 boyut etkisi oluşmuştur. Amblemin üzüm salkımının orta iç kısmında bulunan üzüm tanesi etrafında var olan formu bozmayacak şekilde beyaz bir boşluk yapılmıştır. Bu etki amblemin en alt kısmında bulunan üzüm tanesinin üst kısmında da verilmiştir. Amblemin üst kısmı ise yeşil renkte sap ve üç adet yaprak kısmından oluşmaktadır. Yaprakların sivri uçlu olduğu ve ortasında sivri uç yönünde sivri uçlu negatif boşluk çizgisi olduğu görülmektedir. Logonun üzüm taneleri Görsel 6 ile aynı tarzda ve aynı düzen içerisindedir.

Logonun Grape Juice yazılı logotype yazısını şekilsel açıdan en iyi Görsel 7 uyarlamıştır. Diğer yapay zeka araçları logotype yazısındaki harfleri baş harfi büyük diğer harfleri küçük harflerle yazmıştır.



Çalışma 2.1. Ambalaj Tasarımı: Kader Horasan, 2023

Prompt: "*Bu yüklenen logo tasarımı için bir ambalaj tasarımı oluştur. Ambalaj dikey ve yuvarlak bir cam şişeden oluşsun. Kapağı plastik beyaz çevir aç kapak olsun. İçerisindeki meyve suyu mor renk olsun. Arka planda herhangi bir ayrıntı olmasın ve beyaz renkte olsun. Üzerindeki etiketteki logonun üst kısmında koyu yeşil bant olsun ve içinde de 'Doğanın Meyvede Saklı Hali' yazsın. Logonun alt kısmında mor renkte '%100 Doğal Üzüm Suyu' altında yeşil renkte 'Katkı Maddesi İçermez' altında mor renkte 'Soğuk Sıkım' yanına da 'Geleneksel Tarif' yazsın. Etiketin en alt kısmında '%100 Üzüm - Şeker İlavesiz - Koruyucu İçermez' yazısı sembolleriyle birlikte olsun*." Belirtilen prompt doğrultusunda yapay zeka araçlarının oluşturduğu görsel ambalaj tasarımı örnekleri aşağıda verilmiştir.



Görsel 5.1. ChatGPT (2026)



Görsel 6.1. Gemini (2026)



Görsel 7.1. Ideogram (2026)



Görsel 8.1. Runway (2026)

Çalışma 2.1.'deki ambalaj tasarımına ait prompt girildiğinde ortaya çıkan bulgular aşağıda verilmiştir;

Çalışma 2.1.'deki ambalaj tasarımındaki gibi Görsel 5.1., Görsel 6.1., Görsel 7.1. ve Görsel 8.1.'de Ambalaj dikey ve yuvarlak bir cam şişeden oluştuğu görülmektedir. Kapak incelendiğinde, sadece Görsel 5.1., plastik beyaz çevir aç kapak diğerleri ise metal kapak olarak görülmektedir. Etiketin en alt kısmında '%100 Üzüm-Şeker İlavesiz-Koruyucu İçermez' yazısı sembollerinin birbirinden farklı olduğu görülmektedir. Yapay zeka ile üretilen tüm ambalaj tasarımları kendi içerisinde değerlendirildiğinde, Görsel 5.1. ve Görsel 7.1. etiketinin üzerinde logonun çerçeve alanı kaldırılmıştır. Ambalajdaki etiketin şekli ve şişe formu açısından Görsel 5.1., Görsel 7.1. ve Görsel 8.1. benzerlik göstermektedir.



Çalışma 3. Orijinal Logo Tasarımı: Fatma Çakır, 2024

Çalışma 3'te çilek şeklinden çizgisel olarak tasarlanmış bir logo görülmektedir. Amblemin, yeşil sap ve kırmızı renkte, belirli bir denge ve düzen içerisinde, çizgisel nüanslı helezonik ters kıvrım hareketleriyle çilek formundan oluşturulduğu görülmektedir. Amblem ortalama 45° açıyla sol tarafa eğimli olarak logotype yazısının ölçüsüne hizalanmıştır. Amblemin alt orta kısmına ise 'Strawberry Chocolate' yazılı serifsiz, alt alta iki kelimeden oluşan logotype yazısı konumlandırılmıştır. Logo yeşil, kırmızı ve kahverengi renkte ve çerçevesiz olarak tasarlanmıştır.

Prompt: "*Bir logo tasarımı oluştur. Logo, çilek şeklinde çizgisel olarak helezonik ters kıvrım hareketleri olan stilize bir amblem içersin. Amblem 45° açıyla sol tarafa eğimli ve rengi kırmızı sap kısmı ise yeşil olsun. Logonun altında iki satır olacak şekilde 'Strawberry Chocolate' yazılı serifsiz, alt alta iki kelimeden oluşan logotype yazısı olsun*." Belirtilen prompt doğrultusunda yapay zeka araçlarının oluşturduğu görsel logo örnekleri aşağıda verilmiştir.



Görsel 9. ChatGPT (2026)



Görsel 10. Gemini (2026)



Görsel 11. Ideogram (2026)



Görsel 12. Runway (2026)

Çalışma 3'deki logoya ait prompt girildiğinde ortaya çıkan bulgular aşağıda verilmiştir;

Görsel 9'daki amblemde çilek şeklinde çizgisel olarak helezonik ters kıvrım hareketleri vardır. Fakat bu hareketlerin bir kısmı yarım kalmıştır. Yeşil sap kısmında tarzından farklı olarak keskin hatlara sahip sivri ve köşeli ayrıntılar mevcuttur. Amblem sol tarafa eğimli olarak logotype yazısından daha küçük ölçüde hizalanmıştır.

Görsel 10'daki çilek şeklinde çizgisel olarak helezonik kıvrım hareketleri vardır ve çilek formunda boyutlu kıvrımlı çizgilerin olduğu, bu çizgilerin bazılarının birbirine bağlantısız olduğu görülmektedir. Yeşil sap kısmında tarzından farklı olarak keskin hatlara sahip sivri ve köşeli ayrıntılar mevcuttur. Sap ile çilek şeklinin birleşim kısmının yönü ve sapın bulunduğu yön farklı açılardadır. Amblem sol tarafa eğimli olarak tasarlanmıştır.

Görsel 11'deki çilek şeklinde çizgisel olarak helezonik ters kıvrım hareketleri vardır. Amblemde çilek formunda boyutlu kıvrımlı çizgilerin olduğu görülmektedir. Yeşil sap kısmı, keskin hatlara ve yeşil renk dolgusuna sahiptir. Amblem sağ tarafa eğimli olarak tasarlanmıştır.

Görsel 12'deki çilek şeklinde çizgisel olarak birbirine paralel, spiral şeklinde helezonik kıvrım hareketleri ve bu hareketlerin devamlılığını kesen küçük noktalar vardır. Çilek şeklindeki kıvrımlı çizgilerinin bazılarının birbirine bağlantısız olduğu görülmektedir. Yeşil sap kısmında tarzından farklı olarak keskin hatlara sahip sivri ve köşeli ayrıntılar mevcuttur. Amblem sağ tarafa eğimli olarak tasarlanmıştır.

Logonun Strawberry Chocolate yazılı logotype yazısını şekilsel açıdan en iyi Görsel 9 uyarlamıştır. Diğer yapay zeka araçları logotype yazısındaki harflerin tamamının büyük ya da küçük yazmıştır. Yapay zeka araçlarının oluşturduğu görseller amblem ile logotype arasında oran orantı ve kompozisyon açısından uyumsuzluklar vardır. Görsel 9 ve Görsel 11'de logotype yazısı kırmızı renkte olduğu, Görsel 10 ve Görsel 12'de ise kahverengi olduğu görülmektedir.



Çalışma 3.1. Ambalaj Tasarımı: Fatma Çakır, 2024

Prompt: "*Bu yüklenen logo tasarımı için bir ambalaj tasarımı oluştur. Ambalaj dikdörtgen dikey şekilde beyaz renkte bir çikolata ambalajı olarak tasarlansın. Ambalajın sağ ve sol yanında dikey olarak kırmızı şerit olsun. Ambalajın üst orta kısmında logo olsun. Ambalajın alt orta kısmında kırmızı renkte '%75 COCOA' yazsın. Altında daha küçük oranda kırmızı renkte '%100 NATURAL' yazsın ve altında biraz büyük '120g' yazsın*." Belirtilen prompt doğrultusunda yapay zeka araçlarının oluşturduğu görsel ambalaj tasarımı örnekleri aşağıda verilmiştir.



Görsel 9.1. ChatGPT (2026)



Görsel 10.1. Gemini (2026)



Görsel 11.1. Ideogram (2026)



Görsel 12.1. Runway (2026)

Çalışma 3.1.'deki ambalaj tasarımına ait prompt girildiğinde ortaya çıkan bulgular aşağıda verilmiştir;

Çalışma 3.1.'deki ambalaj tasarımındaki gibi Görsel 9.1., Görsel 10.1., Görsel 11.1. ve Görsel 12.1.'de ambalaj, dikey dikdörtgen şekilde, karton kağıt bir malzemeden ve beyaz renkte, sağ ve sol tarafında kırmızı şeritlerden oluştuğu görülmektedir. Ambalajlardaki kırmızı şeritler incelendiğinde, şerit kalınlıklarının birbirinden farklı olduğu görülmektedir. Sadece Görsel 12.1. deki ambalajın üzerindeki kırmızı şeritlere, mevcut logodaki hareketli helezonik çizgileri yerleştirdiği görülmektedir. Ambalajların en alt kısmında, alt alta konumlandırılmış '%75 Cocoa %100 Natural 120 G' yazısının font, büyüklük küçüklük ve düzen açısından birbirinden farklı olduğu görülmektedir. Yapay zeka ile üretilen tüm ambalaj tasarımları kendi içerisinde değerlendirildiğinde, Görsel 10.1. ambalajının üzerinde yer alan '120 G' yazısının renk olarak diğer ambalaj tasarımlarından farklı olduğu görülmektedir. Görsel 12.1. ambalajının üzerinde bulunan '%75 Cocoa %100 Natural 120 G' yazısının tasarım bütünlüğünü bozacak derecede büyük olduğu ve aynı yazının Görsel 9.1.'de diğer tüm görsellere kıyasla daha sade ve küçük olduğu dikkat çekmektedir.



Çalışma 4. Orijinal Logo Tasarımı: Gülçin Kıvrak, 2025

Çalışma 4'te ahtapot figüründen tasarlanmış bir logo görülmektedir. Amblem, lacivert renkte ahtapot figüründen yapılmıştır. Denge ve düzen içerisinde tasarlanmış olan ahtapot figürünün kıvrımlı kolları illüstrasyon tekniğinden yararlanılarak yapılmıştır. Amblem mavi bir halkanın iç kısmındaki beyaz zemine yerleştirilmiştir. Amblemdeki figür mavi yuvarlak çerçevenin içerisinde beyaz zeminde mavi renkte ve sol tarafa bakar konumdadır. Figürün iki kolu yukarı hareket halinde iken alt kısımda bulunan kolları asimetrik denge içerisinde tasarlanmıştır. Logotype yazısı lacivert halkanın eğimine göre üst orta kısma hizalanmış ve amblemin illüstratif tarzına uygun 'AQUISO' yazılı kıvrımlı bir fontla yazıldığı görülmektedir. Ayrıca markanın isminden ayrı olarak, mavi halkanın alt orta kısmına hizalanmış iki kelimeden oluşan serifsiz 'Frozen Seafood' yazılı bir ayrıntı mevcuttur. Logo mavi ve beyaz renkte tasarlanmıştır.

Prompt: "*Bir logo tasarımı oluştur. Logo, ahtapot figüründen tasarlanmış illüstratif bir amblem içersin. Ahtapot figürü beyaz zeminde mavi renk olsun. Ahtapot figürünün kolları kıvrımlı asimetrik denge ve düzen içerisinde illüstratif bir tarzda görünsün. Ahtapot figürü sol tarafa bakar konumda olsun. Logonun yuvarlak mavi halka şeklinde dış çizgisi olsun. Logotype yazısı mavi halkanın eğimine göre beyaz renkte üst orta kısma hizalanmış ve amblemin illüstratif tarzına uygun 'AQUISO' yazılı kıvrımlı bir fontla yazılsın. Ayrıca markanın isminden ayrı olarak, mavi halkanın alt orta kısmına hizalanmış beyaz renkte iki kelimeden oluşan 'Frozen Seafood' yazılı serifsiz bir yazı mevcut olsun*." Belirtilen prompt doğrultusunda yapay zeka araçlarının oluşturduğu görsel logo örnekleri aşağıda verilmiştir.



Görsel 13. ChatGPT (2026)



Görsel 14. Gemini (2026)



Görsel 15. Ideogram (2026)



Görsel 16. Runway (2026)

Çalışma 4'teki logoya ait prompt girildiğinde ortaya çıkan bulgular aşağıda verilmiştir;

Görsel 13'deki amblemde merkezde konumlanan stilize ahtapot figürü karşıdan simetrik bir denge içerisinde yer almaktadır. Ahtapot figürü illüstratif bir tarzda görünmektedir. Logonun yuvarlak mavi dış çizgisine beyaz zemin alanı taşmıştır.

Görsel 14'deki amblemde merkezde konumlanan stilize ahtapot figürünün kolları kıvrımlı asimetrik denge ve düzen içerisinde illüstratif bir tarzda yer almaktadır. Ahtapot figürü sağ tarafa bakar konumdadır. Ahtapot figürü illüstratif bir tarzda görünmekte ve logonun yuvarlak mavi dış çizgisine ahtapot figürünün kolları gelerek beyaz zemin alanı ile helezonik bir beyaz çizgi alanı oluşturmuştur.

Görsel 15'deki amblemde merkezde konumlanan stilize ahtapot figürü sol tarafa bakar konumda olup sağ yöne eğimli simetrik bir denge içerisinde yer almaktadır. Amblemdeki figürün oval şekillerin (vantuz) büyükten küçüğe doğru statik tekrarı figürde şablon etkisi oluşturmaktadır.

Görsel 16'daki amblemde merkezde konumlanan stilize ahtapot figürü sol tarafa bakar konumda olup asimetrik bir denge içerisinde görünmektedir. Ahtapot figürü illüstratif bir tarzda görünmektedir. Beyaz zemin alanının içerisinde ahtapot figürü kontur çizgilerle verilmiştir. Logonun yuvarlak mavi dış çizgisinin ölçüsü beyaz olan yuvarlak alan ile ölçülü olmadığı görülmektedir. Mavi alanın yukarı kısmının kalın alt kısmının ise daha ince olduğu görülmektedir.

Logotype yazısı mavi halkanın eğimine göre beyaz renkte üst orta kısma hizalanmış 'AQUISO' yazılı kıvrımlı bir fontla yazılmıştır. Yapay zeka ile oluşturulan görsellerdeki fontların birbirine benzer nitelikte olduğu görülmektedir. Görsel 14, Görsel 15 ve Görsel 16'da mavi alanın alt kısmında 'Frozen Seafood' yazılı serifsiz yazısı birbirine benzer niteliktedir.



Çalışma 4.1. Ambalaj Tasarımı: Gülçin Kıvrak, 2025

Prompt: "*Bu yüklenen logo tasarımı için bir ambalaj tasarımı oluştur. Ambalaj dikdörtgen dikey şekilde iki renk olsun. Ambalajın üst kısmı %40'ı beyaz alt kısmı ise %60 oranında koyu mavi renk olsun. Üst beyaz kısmın ortasına logoyu yerleştir. Logonun altına koyu mavi 'FROZEN' yazısı ekle. Ambalajın koyu mavi orta kısmında ise beyaz yuvarlak bir zemin oluştur. Yuvarlak beyaz zemin üstünde alanın dışına taşar şekilde 'pembe tonlarda ahtapot, kırmızı tonda ıstakoz, gri renkte balık, siyah renkte kabuklu midye ve iki somon balığı dilimi' bulunan illüstratif bir kompozisyon oluştur. Beyaz yuvarlak zeminin üst kısmında beyaz renkte 'SEAFOOD' yazsın. Ambalajın sağ alt köşesinde zemindeki mavi tonun açık renginde '500g' yazsın.*" Belirtilen prompt doğrultusunda yapay zeka araçlarının oluşturduğu görsel ambalaj tasarımı örnekleri aşağıda verilmiştir.



Görsel 13.1. ChatGPT (2026)



Görsel 14.1. Gemini (2026)



Görsel 15.1. Ideogram (2026)



Görsel 16.1. Runway (2026)

Çalışma 4.1.'deki ambalaj tasarımına ait prompt girildiğinde ortaya çıkan bulgular aşağıda verilmiştir;

Çalışma 4.1.'deki ambalaj tasarımındaki gibi Görsel 15.1.'deki paket benzerlik göstermektedir. Görsel 13.1., Görsel 14.1. ve Görsel 16.1.'deki ambalajların, dikey dikdörtgen şekilde, karton kağıt bir malzemeden oluştuğu görülmektedir. Ambalajların üzerindeki illüstrasyonlar birbirinin tekrarı niteliğindedir. Bununla birlikte, Görsel 13.1. ve Görsel 15.1.'deki ahtapot, ıstakoz, somon dilimleri, bütün balık ve midye görselleri birebir aynı görünmektedir. Görsel 14.1. ve Görsel 16.1.'deki ahtapot, ıstakoz, somon dilimleri, bütün balık (yönleri farklı) ve midye görselleri birebir aynı görünmektedir. Ayrıca aynı ıstakoz görseli, Görsel 14.1.'de bütün balığın alt kısmındayken, Görsel 16.1.'de bütün balığın üstünde konumlandırılmıştır. Yani tüm görsellerin yüksek oranda birbirine benzediği görülmektedir.



Çalışma 5. Orijinal Logo Tasarımı: Sidregül Morcu, 2025

Çalışma 5'te tilki figüründen tasarlanmış bir logo görülmektedir. Logo, turuncu dikey dikdörtgen ve köşeleri oval şekilde tasarlanmıştır. Logodaki amblem tilki figüründen oluşmaktadır ve turuncu dikdörtgen köşeleri oval şeklin içerisinde yer almaktadır. Amblemin, turuncu dikey şekildeki zeminde, beyaz renkte, keskin çizgisel nüanslı illüstratif çizim şeklinde olduğu görülmektedir. Amblem, tilki figürünün kafa kısmı ve kuyruğuyla tasarlanmıştır. Tilki figürünün kafası, sol tarafa dik bir şekilde bakar konumdadır. Tilki figürünün kuyruğu, sağdan sola doğru kafanın üst kısmına kıvrılmış şekildedir. Turuncu dikdörtgen şeklin içerisinde, amblemin alt orta kısmında, beyaz renkte 'FOXUS' yazılı serifsiz logotype yazısı konumlandırılmıştır. Logo turuncu ve beyaz renkte tasarlanmıştır.

Prompt: "*Bir logo tasarımı oluştur. Logo, turuncu dikey dikdörtgen şekil olarak sadece köşeleri oval olacak şekilde tasarlansın. Logo, tilki figüründen tasarlanmış illüstratif bir amblem içersin. Amblem, turuncu dikey şekildeki zeminde, beyaz renkte, keskin çizgisel nüanslı illüstratif çizim şeklinde olsun. Amblem, tilki figürünün kafa kısmı ve kuyruğuyla tasarlansın. Tilki figürünün kafası, sol tarafa dik bir şekilde bakar konumda olsun. Tilki figürünün kuyruğu, sağdan sola doğru kafanın üst kısmına kıvrılsın. Turuncu dikdörtgen şeklin içerisinde, amblemin alt orta kısmında, beyaz renkte serifsiz bir fontla 'FOXUS' yazsın.*" Belirtilen prompt doğrultusunda yapay zeka araçlarının oluşturduğu görsel logo örnekleri aşağıda verilmiştir.



Görsel 17. ChatGPT (2026)



Görsel 18. Gemini (2026)



Görsel 19. Ideogram (2026)



Görsel 20. Runway (2026)

Çalışma 5'teki logoya ait prompt girildiğinde ortaya çıkan bulgular aşağıda verilmiştir;

Görsel 17'deki amblemde tilki şekli, turuncu dikey dikdörtgen bir zemine konumlandırılmıştır. Tilkinin kafası ve kuyruğu, keskin ve kıvrımlı çizgilerle ve negatif pozitif etki verilerek ifade edilmiştir. Tilki figürünün kafası, sol tarafa dik bir şekilde bakar konumdadır ve kuyruğu, sağdan sola doğru kafanın üst kısmına kıvrılmıştır. Amblem dikey turuncu dikdörtgen şeklin içerisinde, alt orta kısmında bulunan logotype yazısının üst kısmında ortalı olarak hizalanmıştır.

Görsel 18'deki amblemde tilki şekli, turuncu, dikey, köşeleri oval bir dikdörtgen bir zemine konumlandırılmıştır. Tilkinin kafası ve kuyruğu, keskin ve kıvrımlı çizgilerle ve negatif pozitif etki verilerek ifade edilmiştir. Tilki figürünün kafası, sol tarafa aşağı eğimli bir şekilde bakar konumdadır ve kuyruğu, sağdan sola doğru kafanın üst kısmına kıvrılmış şekildedir. Amblem dikey turuncu dikdörtgen şeklin içerisinde, alt orta kısmında bulunan logotype yazısının üst kısmında ortalı olarak hizalanmıştır.

Görsel 19'daki amblemde tilki şekli, turuncu, dikey, köşeleri oval bir dikdörtgen zemine konumlandırılmıştır. Tilkinin kafası ve kuyruğu, keskin ve kıvrımlı çizgilerle ve negatif pozitif etki verilerek ifade edilmiştir. Tilki figürünün kafası, doğrudan karşıya izleyiciye bakar konumdadır ve kuyruğu, soldan sağa doğru kafanın üst kısmına paralel biçimde uzatılmıştır. Amblem hem dikey turuncu dikdörtgen şekle

hem de logotype yazısının üst kısmına ortalı olarak hizalanmıştır.

Görsel 20'deki amblemde tilki şekli, turuncu, dikey, köşeleri oval bir dikdörtgen zemine konumlandırılarak, dikdörtgenin iç kısmına beyaz renkte ince bir şeritle çerçevelenmiştir. Tilkinin kafası ve kuyruğu, keskin, kıvrımlı ve illüstratif çizgilerle, negatif pozitif etki verilerek daha detaylı ifade edilmiştir. Tilki figürünün kafası, sol tarafa aşağı yönde eğimli bir şekilde bakar konumdadır ve kuyruğu, sağdan sola doğru kafanın üst kısmına kıvrılmış ancak bu kıvrım kulak hizasında bitirilmiş şekildedir. Bu etki önde arkada hissi uyandırmıştır. Amblem dikey turuncu dikdörtgen şeklin içerisinde, alt orta kısmında bulunan logotype yazısının üst kısmında ortalı olarak hizalanmıştır.

Logonun FOXUS yazılı logotype yazısıyla, yapay zeka araçlarının görsellerinde şekilsel açıdan benzerlik olmadığı yapay zeka araçlarının oluşturduğu görsellerin logotype açısından birebir aynı yazı font ailesiyle yazıldığı görülmektedir. Yapay zeka araçlarının oluşturduğu görseller amblem ile logotype arasında yön açısından uyumsuzluklar vardır. Görsel 19'da amblemin kafa ve kuyruk yönünün diğer görsellere göre farklı olduğu görülmektedir. Görsel 17, Görsel 18 ve Görsel 20'de tilki figürünün kafa eğimleri ve anatomik yapıları birbirine benzer nitelikte olduğu görülmektedir.



Çalışma 5.1. Ambalaj Tasarımı: Sidregül Morcu, 2025

Prompt: "*Bu yüklenen logo tasarımı için bir ambalaj tasarımı oluştur. Ambalaj dikdörtgen dikey şekilde beyaz renk olsun. Ambalajın üst kısmı kilit sistemli olsun, alt kısmı ise köşeli, geniş ve eni olan bir formda olsun. Ambalaj beyaz renkte filtre kahve ambalajı olsun. Ambalajın sol yan tarafı görülsün ve sol yanı turuncu renkte olsun. Ambalajın ortasına logoyu yerleştir. Logonun altına turuncu renkte orta büyüklükte 'Coffee' yazısı ekle, bu yazının biraz altına da küçük turuncu bir kare içerisine alt alta olacak şekilde '%100 Arabica' yazısı yerleştir. Ambalajın sol alt köşesinde turuncu renkte küçük ebatta '500g' yazsın.*" Belirtilen prompt doğrultusunda yapay zeka araçlarının oluşturduğu görsel ambalaj tasarımı örnekleri aşağıda verilmiştir.



Görsel 17.1. ChatGPT (2026).



Görsel 18.1. Gemini (2026)



Görsel 19.1. Ideogram (2026)



Görsel 20.1. Runway (2026)

Çalışma 5.1.'deki ambalaj tasarımına ait prompt girildiğinde ortaya çıkan bulgular aşağıda verilmiştir;

Çalışma 5.1.'deki ambalaj tasarımındaki gibi Görsel 17.1., Görsel 18.1., Görsel 19.1. ve Görsel 20.1.'de ambalajın, dikey ve dikdörtgen bir kilitli paketten oluştuğu görülmektedir. Görsel 18.1. ve Görsel 20.1.'de ambalajın en alt kısmında '%100 Arabica' yazısının diğerlerinden farklı olarak beyaz zeminde yazıldığı ve sadece Görsel 18.1.'de % sembolünün yazının sağ tarafında yer aldığı görülmektedir. Yapay zeka ile üretilen tüm ambalaj tasarımları kendi içerisinde değerlendirildiğinde, Görsel 20.1.'de ambalajın üzerindeki logonun çerçeve alanı beyaz ince bir şeritle ve ambalajın alt kısmı kalın bir turuncu şeritle çerçevelenmiştir. Genel anlamda, Görsel 17.1., Görsel 18.1., Görsel 19.1. ve Görsel 20.1.'deki ambalajlar oldukça benzerlik göstermektedir. Sadece Görsel 17.1.'in açısı farklı oluşturulmuştur.

IV. SONUÇ, TARTIŞMA VE ÖNERİLER

Çalışma 1'e ait sonuçlar;

Araştırma kapsamında sunulan promptlar doğrultusunda üretilen görseller betimsel olarak analiz edildiğinde, ortaya çıkan görsellerin orijinal tasarım olan Çalışma 1 ile biçimsel bir benzerlik taşımadığı, yapay zeka çıktısı olan Görsel 1, Görsel 2, Görsel 3 ve Görsel 4'ün kendi içerisinde belirgin çizgileri olduğu, çiçek yapısında taklitler barındırdığı ve şekil-biçim açısından benzer özellikler gösterdiği sonucuna ulaşılmıştır.

Çalışma 1.1'e ait sonuçlar;

Ambalaj tasarımı mockup uyarlamasına yönelik prompt girildiğinde yapay zeka araçlarının ürettiği görsellerin sonucu olarak; ortaya çıkan ambalaj tasarımı görselleri Çalışma 1.1.'deki Mockup uyarlamasına göre görsel açıdan oldukça benzerlik taşıdığı, Görsel 1.1. ve Görsel 3.1.'de verilen yön doğruyken, Görsel 2.1. ve Görsel 4.1.'de ise yönün yanlış verildiği görülmüştür. Mockup uyarlamasına göre görüntü açısından şişe, kapak, açı duruşu, renk ve logonun konumu açısından birebir benzerlik gösterdiği sonucuna ulaşılmıştır.

Çalışma 2'ye ait sonuçlar;

Araştırma kapsamında sunulan promptlar doğrultusunda üretilen görseller betimsel olarak analiz edildiğinde, ortaya çıkan görsellerin orijinal tasarım olan Çalışma 2 ile biçimsel bir benzerlik taşımadığı, yapay zeka çıktısı olan Görsel 5 ve Görsel 7 ile üzüm salkımı şeklinin benzerlik taşıdığı sonucuna ulaşılmıştır. Logonun Grape Juice yazılı logotype yazısını şekilsel açıdan en iyi Görsel 7 uyarlamıştır. Fakat doğru rengi oluşturmamıştır. Şekil olarak dikdörtgen dikey çerçeve yerine etrafına kare çerçeve oluşturmuştur. Yapay zeka araçlarının üretmiş olduğu görseller arasında Görsel 5

ile Görsel 7'nin Görsel 6 ile de Görsel 8'in şekil ve tarz açısından benzer özellikler gösterdiği sonucuna ulaşılmıştır. Yapay zeka çıktısı olan her dört görselin kendi içlerinde birbirine benzer nitelikte belirgin çizgileri olduğu, üzüm salkımı yapısında taklitler barındırdığı ve şekil-biçim açısından benzerlikler gösterdiği sonucuna ulaşılmıştır.

Çalışma 2.1'e ait sonuçlar;

Ambalaj tasarımı mockup uyarlamasına yönelik prompt girildiğinde yapay zeka araçlarının ürettiği görsellerin sonucu olarak; Görsel 5.1., Görsel 6.1., Görsel 7.1. ve Görsel 8.1.'de ambalaj dikey ve yuvarlak bir cam şişeden oluşmuş, kapak sadece Görsel 5.1., plastik beyaz çevir aç kapak diğerleri ise metal kapak olarak oluşturulmuştur. Etiketin en alt kısmında '%100 Üzüm-Şeker İlavesiz-Koruyucu İçermez' yazısı sembollerinin birbirinden farklı olduğu görülmüştür. Mockup uyarlamasına göre görüntü açısından Görsel 5.1. ve Görsel 7.1. etiketinin üzerinde logonun çerçeve alanı olmadığı ve ambalajdaki etiketin şekli ve şişe formu açısından Görsel 5.1., Görsel 7.1. ve Görsel 8.1. oldukça benzerlik gösterdiği sonucuna ulaşılmıştır.

Çalışma 3'e ait sonuçlar;

Araştırma kapsamında sunulan promptlar doğrultusunda üretilen görseller betimsel olarak analiz edildiğinde, ortaya çıkan görsellerin orijinal tasarım olan Çalışma 3 ile biçimsel bir benzerlik taşımadığı sonucuna ulaşılmıştır. Logonun Strawberry Chocolate yazılı logotype yazısının şekilsel açıdan en iyi Görsel 9'da uyarlandığı, diğer yapay zeka araçlarının logotype yazısındaki harflerin tamamının büyük ya da küçük yazıldığı, genel olarak tüm araçların tasarım görsellerinde amblem ile logotype arasında oran orantı ve kompozisyon açısından uyumsuzluklar olduğu, Görsel 9 ve Görsel 11'de logotype yazısının kırmızı renkte, Görsel 10 ve Görsel 12'de ise kahverengi olduğu sonucuna ulaşılmıştır. Görsel 10, Görsel 11 ve Görsel 12'nin yeşil sap kısmının şekil ve tarz açısından benzer özellikler gösterdiği, yapay zeka çıktısı olan her dört görselin kendi içlerinde birbirine benzer nitelikte belirgin çizgileri olduğu, helezonik çizgileri birbirinden farklı olsa da bütüne bakıldığında şekil-biçim açısından benzerlikler gösterdiği sonucuna ulaşılmıştır.

Çalışma 3.1'e ait sonuçlar;

Ambalaj tasarımı mockup uyarlamasına yönelik prompt girildiğinde yapay zeka araçlarının ürettiği görsellerin sonucu olarak; ortaya çıkan ambalaj tasarımı görsellerinin hepsinin mockup uyarlamasına göre görsel açıdan oldukça benzerlik taşıdığı sonucuna ulaşılmıştır. Görsel 9.1., Görsel 10.1., Görsel 11.1. ve Görsel 12.1.'de ambalajın, dikey dikdörtgen şekilde, karton kağıt bir malzemeden ve beyaz renkte, sağ ve sol tarafında kırmızı şeritlerden oluştuğu sonucuna ulaşılmıştır. Ambalajlardaki kırmızı şeritlerin kalınlıklarının birbirinden farklı olduğu ve sadece Görsel 12.1. deki ambalajın üzerindeki kırmızı şeritlere, mevcut logodaki hareketli helezonik çizgilerin yerleştirdiği sonucuna ulaşılmıştır. Ambalajların en alt kısmında, alt alta konumlandırılmış '%75 Cocoa, %100 Natural, 120g' yazısının font, büyüklük küçüklük ve düzen açısından birbirinden farklı olduğu, Görsel 10.1. ambalajın üzerinde yer alan '120g' yazısı renk açısından, Görsel 12.1. ambalajının üzerindeki '%75 Cocoa, %100 Natural, 120g' yazısı font büyüklüğü açısından ve Görsel 9.1.'de ise aynı yazının sadelik ve font küçüklüğü açısından oran orantı, görsel hiyerarşi, renk bütünlüğü gibi konularda eksik olduğu sonucuna ulaşılmıştır.

Çalışma 4'e ait sonuçlar;

Araştırma kapsamında sunulan promptlar doğrultusunda üretilen görseller betimsel olarak analiz edildiğinde, ortaya çıkan görsellerin orijinal tasarım olan Çalışma 4 ile genel anlamda biçimsel benzerliklerin ve ayrıntıda farklılaşmaların olduğu sonucuna ulaşılmıştır. Logotype yazısının, mavi halkanın eğimine göre beyaz renkte üst orta kısma hizalanmış 'AQUISO' yazılı kıvrımlı bir fontla yazıldığı, yapay zeka ile oluşturulan görsellerdeki fontların birbirine benzer nitelikte olduğu ve Görsel 14, Görsel 15 ve Görsel 16'da mavi alanın alt kısmında 'Frozen Seafood' yazılı serifsiz yazısının birbirine benzer nitelikte olduğu sonucuna ulaşılmıştır.

Çalışma 4.1'e ait sonuçlar;

Ambalaj tasarımı mockup uyarlamasına yönelik prompt girildiğinde yapay zeka araçlarının ürettiği görsellerin sonucu olarak; ortaya çıkan ambalaj tasarımı görsellerinin hepsinin mockup uyarlamasına göre görsel açıdan benzerlik taşıdığı sonucuna ulaşılmıştır. Görsel 15.1.'deki paketin Çalışma 4.1.'deki ambalaj tasarımıyla benzerlik gösterdiği, Görsel 13.1., Görsel 14.1. ve Görsel 16.1.'deki ambalajların, dikey dikdörtgen şekilde, karton kağıt bir malzemeden oluştuğu, ambalajların üzerindeki illüstrasyonlar birbirinin tekrarı niteliğinde olduğu sonucuna ulaşılmıştır. Bununla birlikte, Görsel 13.1. ile Görsel 15.1.'deki ahtapot, ıstakoz, somon dilimleri, bütün balık ve midye görsellerinin birebir aynı göründüğü ve Görsel 14.1. ile Görsel 16.1.'deki ahtapot, ıstakoz, somon dilimleri, bütün balık (yönleri farklı) ve midye görsellerinin birebir aynı göründüğü sonucuna ulaşılmıştır. Ayrıca aynı ıstakoz görselinin, Görsel 14.1.'de bütün balığın alt kısmındayken, Görsel 16.1.'de bütün balığın üstünde konumlandırıldığı, yani tüm görsellerin yüksek oranda birbirine benzediği sonucuna ulaşılmıştır.

Çalışma 5'e ait sonuçlar;

Araştırma kapsamında sunulan promptlar doğrultusunda üretilen görseller betimsel olarak analiz edildiğinde, ortaya çıkan görsellerin orijinal tasarım olan Çalışma 5 ile biçimsel bir benzerlik taşımadığı sonucuna ulaşılmıştır. Görsel 17, Görsel 18, Görsel 19 ve Görsel 20'deki amblemde tilki şekli, turuncu dikey dikdörtgen bir zemine konumlandırılmış, tilki figürünün kafası ve kuyruğu, keskin ve kıvrımlı çizgilerle verildiği, beyaz ve turuncu renk ile Görsel 17, Görsel 18 ve Görsel 20'de negatif pozitif etki verilerek ifade edildiği, Görsel 19'da tilki figürünün kafası, doğrudan karşıya izleyiciye bakar konumda olduğu ve Görsel 17, Görsel 18 ve Görsel 20'de sol tarafa aşağı yönde eğimli bir şekilde bakar konumda olduğu sonucuna ulaşılmıştır. Logonun FOXUS yazılı logotype yazısıyla, yapay zeka araçlarının görsellerinde şekilsel açıdan benzerlik olmadığı yapay zeka araçlarının oluşturduğu görsellerin logotype açısından birebir aynı yazı font ailesiyle yazıldığı, yapay zeka araçlarının oluşturduğu görsellerde amblem ile logotype arasında yön açısından uyumsuzluklar olduğu, Görsel 17, Görsel 18 ve Görsel 20'de tilki figürünün kafa eğimleri ve anatomik yapıları birbirine benzer nitelikte olduğu sonucuna ulaşılmıştır.

Çalışma 5.1'e ait sonuçlar;

Ambalaj tasarımı mockup uyarlamasına yönelik prompt girildiğinde yapay zeka araçlarının ürettiği görsellerin sonucu olarak; ortaya çıkan ambalaj tasarımı görselleri mockup uyarlamasına göre görsel açıdan oldukça benzerlik taşıdığı sonucuna ulaşılmıştır. Görsel 17.1., Görsel 18.1., Görsel 19.1. ve Görsel 20.1.'deki ambalajların Çalışma 5.1.'deki ambalaj tasarımına neredeyse birebir benzediği, ayrıca tüm yapay zeka görsellerinin kendi içinde de ambalajın üzerindeki zemin rengi ve sembollere ait yerlerin farklılaşması dışında birbirlerine yüksek oranda birebir benzediği sonucuna ulaşılmıştır.

Yapay zekanın, internet ortamındaki verilerin birleşiminden ortaya çıkan ürünler sunduğu, araştırmanın bulgular bölümünde kanıtlanmaktadır. Bu bağlamda, eğer internet ortamında kişiye ait bir bilgi ya da veri yoksa yapay zekanın kullanıcıya sunabileceği bir veri yoktur. İstenilen tasarım her ne olursa olsun yapay zeka araçlarının, bu tasarım ürününü internet ortamındaki verileri birleştirerek görselleştireceği ve kullanıcıya sunacağı bir gerçektir. Bu durum bulgular bölümünde verilen orijinal tasarım ürünleri ile yapay zeka araçlarının ürettiği görseller arasındaki farklılıktan ve ayrıca söz konusu yapay zeka araçlarının ürettiği görsellerin kendi içerisindeki yüksek orandaki benzeşmeden net bir şekilde görülmektedir. Söz konusu benzeşimin yüksek oranda olmasının sebebi ise yapay zeka araçlarının zaten var olan hazır verileri çekerek o görselleri birleştirerek tasarlamasıdır.Bu da tasarım alanında çok ciddi etik sorunları beraberinde getirecektir. Bu nedenle yapay zekadan yararlanırken oldukça dikkat edilmeli etik ihlaline karşı önlem alınmalıdır. Bir tasarımcı sunumunu yaparken yapay zekadan yararlanabilir. Ancak bir amaç olarak değil sadece araç olarak kullanmalıdır.

Yapay zekadan görsel oluşturulduğunda girilen promptu bir tasarımcıymış gibi algılamamak gerekir. Yapay zeka mekanik bir araç iken tasarımcı olmak yaratıcı düşünce, görsel algı, estetik bilinç ve tasarım becerisi gerektirmektedir. Dolayısıyla yapay zeka algoritmasının insanoğlunun yönlendirdiği bir araç olduğu unutulmamalıdır. Eğitimcilerin bu konuda bilinçli olması ve bu bilinci öğrencilere de kazandırması gerekliliği önemli bir husustur. Ergen (2022)'in de belirttiği üzere, yapay zekayla birlikte, el becerileri daha az değerli hale gelmiş ancak aynı zamanda yaratıcı bir zihnin önemi de artmış olacaktır [11].

Sanat veya tasarım açısından yapay zeka üretimi bir eser ya da tasarım ürünü olarak görünse de aslında komut ve düşünceye dayalı bir örnek sunu niteliğindedir. Özgünlük ve yaratıcı düşünce sonucu duygu ve düşüncelerle ortaya çıkartılmış özgün bir tasarım anlayışı ne yazık ki yoktur. Bunun için her zaman insan düşüncesine ihtiyaç vardır. Bu da çalışmanın bulgular bölümünde, hem logo tasarımlarında hem de ambalaj tasarımlarında ve illüstrasyon çizimlerinde net bir şekilde kanıtlanmaktadır.

Çalışmanın bu başlığı altında ilgili alan yazın incelenerek, bu araştırma ile benzerlik gösteren araştırmalara yer verilmiştir. Bu kapsamda, tespit edilen benzer çalışmalar ayrıntılı olarak tartışılmıştır. Bu araştırmalar aşağıda yer almaktadır.

Hwang ve Wu (2025)'nun 'Graphic Design Education in the Era of Text-to-Image Generation: Transitioning to Contents Creator/Metin-Görsel Üretimi Çağında Grafik Tasarım Eğitimi: İçerik Üreticiliğine Geçiş' başlıklı makale çalışmasında; üretken yapay zekanın ilerlemesi, çeşitli endüstrilerde ve akademik alanlarda yenilikçi fırsatlar ve yeni zorluklar sunduğu belirtilerek, özellikle metinden görsel üretebilen üretken yapay zekadaki son gelişmelerin, grafik tasarım eğitimi alanında yeni zorluklar ortaya çıkardığı ifade edilmektedir. Çalışmada, 'Görsel İletişim Tasarımı' dersinde yürütülen atölye çalışmalarına dayanarak üretken yapay zekayı kullanan grafik tasarım eğitimi yöntemlerinin ele alındığı görülmektedir. Çalışmada öğrenciler, 'Technology Pedagogical Content Knowledge/Teknoloji Pedagoji İçerik Bilgisi (TPCACK)' modeline dayalı olarak Midjourney ve DALL-E gibi üretken yapay zeka araçlarını kullanarak Avrupa tasarım tarihini anlatan posterler oluşturmuşlardır. Makale çalışmasındaki araştırmacılar, vaka çalışmaları ve anket sonuçlarının analizi yoluyla, üretken yapay zeka çağında grafik tasarım eğitiminin, geleneksel eğitimden farklı eğitim hedefleri ve yöntemleri gerektirdiğini doğrulamışlardır. Makale öğrencilerin, uygun anahtar kelimeler kullanarak görselleri doğru şekilde aktarabilmeleri için yapay zeka görsel okuryazarlığı becerilerine ihtiyaçları olduğunu belirtir. Ayrıca eğitimciler, grafik hikâye anlatımı yoluyla yaratıcı anahtar kelimeler türetmeleri için fırsatlar sunmasını ve bilim, sanat ve teknolojiyi birleştiren içerik üreticileri olarak bir vizyon sunması gerektiğini belirtir. Araştırma, üretken yapay zeka çağında grafik tasarımcıların rolleri ve görsel tasarım eğitiminin yönü konusunda kılavuzlar sunmaktadır [12].

Makale çalışmasında yapay zeka araçları kullanılarak grafik tasarım eğitimi alan öğrencilere yapay zeka desteğiyle tasarım oluşturmaları istenmiştir. Bu çalışmada ise öğrencilerin üretmiş olduğu orijinal grafik tasarımı ürünlerinden yola çıkarak, yapay zeka araçlarına prompt girilerek üretilen orijinal tasarım ile yapay zeka araçlarının ürettiği görseller karşılaştırılmıştır.

Muji, Svensson ve Faraon (2023)'a ait 'Engaging With Artificial Intelligence in Graphic Design Education/Grafik Tasarım Eğitiminde Yapay Zekayla Etkileşim Kurmak' başlıklı makale çalışmasında; yüksek öğrenimdeki grafik tasarım derslerinde yapay zekanın potansiyel kullanımı incelenmiştir. Çalışmada öğrencilerin grafik tasarım için altı yapay zeka kaynağını kullanma konusundaki öğrenme deneyimlerini değerlendirmek için nitel görüşmeler yapılmıştır: Limbiks, ChatGPT, Looka, Pagegenie, Khroma ve DALL-E. Sonuçlar, yapay zekanın öğrencilerin grafik tasarım eğitiminde öğrenme deneyimlerini geliştirebileceğini göstermektedir. Ayrıca, öğrencilerin yapay zeka tarafından üretilen bilgileri değerlendirirken eleştirel düşünme becerilerini geliştirmelerine ihtiyaç duyulduğu ortaya çıkmıştır. Makale bulgularının, yapay zeka destekli kaynakların yükseköğretimde öğrenme amaçlı nasıl kullanılabileceğine dair tartışmalar için bir temel oluşturduğu görülmektedir [13].

Çalışmanın bu araştırmayla benzerlik gösterdiği tarafı, farklı yapay zeka araçları kullanılarak grafik tasarım eğitimi alan öğrencilere yapay zeka desteğiyle tasarım oluşturmaları istenmesidir. ChatGPT adlı yapay zeka aracının, her iki çalışmada da kullanıldığı, diğer araçlarda benzerlik olmadığı görülmektedir.

Yapay zekanın problemlerine değinen Nabiyev (2021) ise bilgisayar yaratıcılığı için genellikle insana özgü zihinsel durumların modellenmesinin amaçlandığını belirterek, fotoğrafın resme dönüştürülmesini, resmin fotoğrafa çevrilmesini, ifadeden çizimin yapılmasını, resim eserlerinin belli bir ressama özgü şekilde çizdirilmesini, bilgisayarlı müzik besteleme, şiirlerin, hikaye ve romanların yazılmasını, bu alanın problemleri olarak ele alır. Ayrıca örnek verdiği durumlarda bilgisayar yaratıcılığında belirli ilerlemeler söz konusu olsa da gerçek yaratıcılıktan çok uzakta olduğunu vurgular [3].

Yapay zekanın üretmiş olduğu grafiksel görsellerin internet ortamındaki görsellerle birleştirilerek kullanıcıya veya izleyiciye sunulması ciddi sorunlar oluşturabilir. Yapay zekanın üretmiş olduğu görseller tasarım olarak nitelendirilemez. Bu durum ancak fikir sunma açısından fayda sağlayabilir. Örneğin, özgün bir tasarımı sadece sunuş amacıyla ambalaj tasarımında nasıl duracağını görmek için mock up kullanılabilir. Bu araştırma da bu sebeple, söz konusu durumun açıklığa kavuşması ve bilinçli grafik tasarımcılar yetiştirmek ihtiyacına cevap niteliğini taşımaktadır.

KAYNAKÇA

[1] N. J. Nilsson, Yapay Zekâ, Geçmişi ve Geleceği. M. Doğan Çev., İstanbul: Boğaziçi Üniversitesi Yayınevi, 2018.

[2] S. J. Russell, & P. Norvig, Artificial Intelligence a Modern Approach. third edition, New Jersey: Pearson Education Inc, 2010.

[3] V. Nabiyev, Yapay Zeka: Makine Öğrenmesi, Stratejili Oyunlar, Örüntü Tanıma, Doğal Dil İşleme. Genişletilmiş ve Güncelleştirilmiş 6. baskı, Ankara: Seçkin Yayıncılık, 2021.

[4] R. Başaran, & Y. Yeşilbaş Özenç, "Bilimsel araştırma sürecinde yapay zeka araçlarının kullanımı," Uluslararası Eğitimde Mükemmellik Arayışı Dergisi (UEMAD), *4*(1), s.35-53, 2024.

[5] E. Karaata, "Usage of artificial intelligence in today's graphic design," Online Journal of Art and Design, *6*(4), pp.183-198, 2018.

[6] S. B. Merriam, "Nitel Araştırma Desen ve Uygulama İçin Bir Rehber," S. Turan, Çeviri Ed. 2. Bölüm, F. Koçak Canbaz ve M. Öz, Bölüm Çev., Ankara: Nobel Akademik Yayıncılık, 2023.

[7] J. W. Creswell, "Araştırma Deseni Nitel, Nicel ve Karma Yöntem Yaklaşımları," S. B. DEMİR, Çeviri Ed. 9. Bölüm, Y. Dede, Bölüm Çev., 3. Baskı, Ankara: Eğiten Kitap Yayıncılık, 2017.

[8] S. B. Merriam, "Nitel Araştırma Desen ve Uygulama İçin Bir Rehber," S. Turan, Çeviri Ed. 7. Bölüm, S. Turan ve H. Özen, Bölüm Çev., Ankara: Nobel Akademik Yayıncılık, 2023.

[9] A. Yıldırım, & H. Şimşek, "Sosyal Bilimlerde Nitel Araştırma Yöntemleri," 10. baskı, Ankara: Seçkin Yayıncılık, 2016.

[10] B. Kıral, "Nitel bir veri analizi yöntemi olarak doküman analizi," Siirt Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, *8*(15), s.170-189, 2020.

[11] İ. Ergen, "Yapay Zeka ile Gelişen Tasarım, Grafik ve Oyun Tasarımı, Otonom Sanat, Metaverse ve NFT Olasılıklarını Keşfetmek," İstanbul: Efe Akademi Yayınları, 2022.

[12] Y. Hwang, & Y. Wu, "Graphic design education in the era of text-to-ımage generation: transitioning to contents creator," International Journal of Art & Design Education, *44*(1), pp.239–253, 2025. https://doi.org/10.1111/jade.12558

[13] S. Muji, E. Svensson, & M. Faraon, "Engaging with artificial intelligence in graphic design education," 2023 5th International Workshop on Artificial Intelligence and Education (WAIE), pp.31–37, 2023. https://doi.org/10.1109/WAIE60568.2023.00013

Görsel Kaynaklar

Çalışma 1: Hamdican Yanıkoğlu, Logo Tasarımı, 2024.

Çalışma 1.1: https://unblast.com/download/2002/ adresinden Hamdican Yanıkoğlu tarafından ambalaj tasarımı için Mockup uyarlaması oluşturulmuştur.

Çalışma 2: Kader Horasan, Logo Tasarımı, 2023.

Çalışma 2.1: https://graflick.com/grape-juice-bottle-mockup-psd-free-download/ adresinden Kader Horasan tarafından ambalaj tasarımı için Mockup uyarlaması oluşturulmuştur.

Çalışma 3: Fatma Çakır, Logo Tasarımı, 2024.

Çalışma 3.1: https://mockups-design.com/free-chocolate-mockup/ adresinden Fatma Çakır tarafından ambalaj tasarımı için Mockup uyarlaması oluşturulmuştur.

Çalışma 4: Gülçin Kıvrak, Logo Tasarımı, 2025.

Çalışma 4.1: https://unblast.com/front-view-plastic-pouch-mockup/ adresinden Gülçin Kıvrak tarafından ambalaj tasarımı için Mockup uyarlaması oluşturulmuştur.

Çalışma 5: Sidregül Morcu, Logo Tasarımı, 2025.

Çalışma 5.1: https://goodmockups.com/free-coffe-bag-standing-pouch-mockup-psd-3/ adresinden Sidregül Morcu tarafından ambalaj tasarımı için Mockup uyarlaması oluşturulmuştur.

Yapay Zeka Üretimi Görsel Kaynaklar

Görsel 1: OpenAI. (2026). AI-generated image created with ChatGPT/DALL·E, [Create a logo design. The logo consists of a stylized emblem in the form of a half flower with 9 (petal) leaves arranged side by side. Let the color be dark navy blue. Below the logo, there should be a single-line serifed logotype consisting of two words side by side, reading 'Green Essence'.] OpenAI. https://chatgpt.com

Görsel 1.1: OpenAI. (2026). AI-generated image created with ChatGPT/DALL·E [Create a packaging design for this uploaded logo design. The packaging should be a square glass transparent perfume bottle. The cap should be transparent with square corners. The perfume liquid should be turquoise in color. The perfume bottle should be tilted to the right at a 45° angle. There should be no details in the background and it should be white.] OpenAI. https://chatgpt.com

Görsel 2: Google Gemini. (2026). AI-generated image created with Google Gemini. [Create a logo design. The logo consists of a stylized emblem in the form of a half flower with 9 (petal) leaves arranged side by side. Let the color be dark navy blue. Below the logo, there should be a single-line serifed logotype consisting of two words side by side, reading 'Green Essence'.] Google Gemini https://gemini.google.com/

Görsel 2.1: Google Gemini. (2026). AI-generated image created with Google Gemini. [Create a packaging design for this uploaded logo design. The packaging should be a square glass transparent perfume bottle. The cap should be transparent with square corners. The perfume liquid should be turquoise in color. The perfume bottle should be tilted to the right at a 45° angle. There should be no details in the background and it should be white.] Google Gemini https://gemini.google.com/

Görsel 3: Ideogram.ai. (2026). AI-generated image created with ideogram.ai. [Create a logo design. The logo consists of a stylized emblem in the form of a half flower with 9 (petal) leaves arranged side by side. Let the color be dark navy blue. Below the logo, there should be a single-line serifed logotype consisting of two words side by side, reading 'Green Essence'.] Ideogram.ai. https://ideogram.ai/t/

Görsel 3.1: Ideogram.ai. (2026). AI-generated image created with ideogram.ai. [Create a packaging design for this uploaded logo design. The packaging should be a square glass transparent perfume bottle. The cap should be transparent with square corners. The perfume liquid should be turquoise in color. The perfume bottle should be tilted to the right at a 45° angle. There should be no details in the background and it should be white.] Ideogram.ai. https://ideogram.ai/t/

Görsel 4: RunwayMl. (2026). AI-generated image created with RunwayMl. [Create a logo design. The logo consists of a stylized emblem in the form of a half flower with 9 (petal) leaves arranged side by side. Let the color be dark navy blue. Below the logo, there should be a single-line serifed logotype consisting of two words side by side, reading 'Green Essence'] RunwayML https://app.runwayml.com/

Görsel 4.1: RunwayMl. (2026). AI-generated image created with RunwayMl. [Create a packaging design for this uploaded logo design. The packaging should be a square glass transparent perfume bottle. The cap should be transparent with square corners. The perfume liquid should be turquoise in color. The perfume bottle should be tilted to the right at a 45° angle. There should be no details in the background and it should be white.] RunwayML https://app.runwayml.com/

Görsel 5: OpenAI. (2026). AI-generated image created with ChatGPT/DALL·E, [Create a logo design for me. The logo features a stylized emblem with six grape berries on a stem. Dark purple should be used for the stem, and green for the grape berries. Below the logo, there should be a single-line logo text consisting of two side-by-side words, 'Grape Juice', written in green sans-serif font. Around the logo, there should be a green square frame on a white background.] OpenAI. https://chatgpt.com

Görsel 5.1: OpenAI. (2026). AI-generated image created with ChatGPT/DALL·E, [Create a packaging design for this uploaded logo design. The packaging should consist of a vertical, round glass bottle. The cap should be a plastic white twist-off cap. The fruit juice inside should be purple. There should be no details on the background and it should be white. The logo on the label should have a dark green band at the top, and it should say 'Nature's Hidden State in Fruit' inside the band. At the bottom of the logo, write '100% Natural Grape Juice' in purple, followed by 'Contains No Additives' in green, then 'Cold Pressed' in purple, and finally 'Traditional Recipe'. The bottom part of the label should include the text '100% Grape-No Added Sugar-No Preservatives' along with the corresponding symbols.] OpenAI. https://chatgpt.com

Görsel 6: Google Gemini. (2026). AI-generated image created with Google Gemini. [Create a logo design for me. The logo features a stylized emblem with six grape berries on a stem. Dark purple should be used for the stem, and green for the grape berries. Below the logo, there should be a single-line logo text consisting of two side-by-side words, 'Grape Juice', written in green sans-serif font. Around the logo, there should be a green square frame on a white background.] Google Gemini https://gemini.google.com/

Görsel 6.1: Google Gemini. (2026). AI-generated image created with Google Gemini [Create a packaging design for this uploaded logo design. The packaging should consist of a vertical, round glass bottle. The cap should be a plastic white twist-off cap. The fruit juice inside should be purple. There should be no details on the background and it should be white. The logo on the label should have a dark green band at the top, and it should say 'Nature's Hidden State in Fruit' inside the band. At the bottom of the logo, write '100% Natural Grape Juice' in purple, followed by 'Contains No Additives' in green, then 'Cold Pressed' in purple, and finally 'Traditional Recipe'. The bottom part of the label should include the text '100% Grape-No Added Sugar-No Preservatives' along with the corresponding symbols.] Google Gemini https://gemini.google.com/

Görsel 7: Ideogram.ai. (2026). AI-generated image created with ideogram.ai. [Create a logo design for me. The logo features a stylized emblem with six grape berries on a stem. Dark purple should be used for the stem, and green for the grape berries. Below the logo, there should be a single-line logo text consisting of two side-by-side words, 'Grape Juice', written in green sans-serif font. Around the logo, there should be a green square frame on a white background.] Ideogram.ai. https://ideogram.ai/t/

Görsel 7.1: Ideogram.ai. (2026). AI-generated image created with ideogram.ai. [Create a packaging design for this uploaded logo design. The packaging should consist of a vertical, round glass bottle. The cap should be a plastic white twist-off cap. The fruit juice inside should be purple. There should be no details on the background and it should be white. The logo on the label should have a dark green band at the top, and it should say 'Nature's Hidden State in Fruit' inside the band. At the bottom of the logo, write '100% Natural Grape Juice' in purple, followed by 'Contains No Additives' in green, then 'Cold Pressed' in purple, and finally 'Traditional Recipe'. The bottom part of the label should include the text '100% Grape-No Added Sugar-No Preservatives' along with the corresponding symbols.] Ideogram.ai. https://ideogram.ai/t/

Görsel 8: RunwayMl. (2026). AI-generated image created with RunwayMl. [Create a logo design for me. The logo features a stylized emblem with six grape berries on a stem. Dark purple should be used for the stem, and green for the grape berries. Below the logo, there should be a single-line logo text consisting of two side-by-side words, 'Grape Juice', written in green sans-serif font. Around the logo, there should be a green square frame on a white background.] RunwayML https://app.runwayml.com/

Görsel 8.1: RunwayMl. (2026). AI-generated image created with RunwayMl. [Create a packaging design for this uploaded logo design. The packaging should consist of a vertical, round glass bottle. The cap should be a plastic white twist-off cap. The fruit juice inside should be purple. There should be no details on the background and it should be white. The logo on the label should have a dark green band at the top, and it should say 'Nature's Hidden State in Fruit' inside the band. At the bottom of the logo, write '100% Natural Grape Juice' in purple, followed by 'Contains No Additives' in green, then 'Cold Pressed' in purple, and finally 'Traditional Recipe'. The bottom part of the label should include the text '100% Grape-No Added Sugar-No Preservatives' along with the corresponding symbols.] RunwayML https://app.runwayml.com/

Görsel 9: OpenAI. (2026). AI-generated image created with ChatGPT/DALL·E, [Create a logo design. The logo features a stylized emblem in the shape of a strawberry with linear, helical, reverse-strip patterns. The emblem is slanted at a 45-degree angle to the left, and its stem should be red while the base should be green. Below the two lines of the logo, let there be a sans-serif text reading 'Çilek Çikolata' as a logotype consisting of two words.] OpenAI. https://chatgpt.com

Görsel 9.1: OpenAI. (2026). AI-generated image created with ChatGPT/DALL·E, [Create a packaging design for this uploaded logo design. The packaging should be designed as a rectangular, vertical white chocolate bar wrapper. There should be a vertical red strip on the right and left sides of the packaging. The logo should be placed in the upper center portion of the packaging. The lower middle part of the packaging should say '75% COCOA' in red. Below that, it should say '100% NATURAL' in a smaller red font, and below that, it should say '120 G' in a slightly larger font.] OpenAI. https://chatgpt.com

Görsel 10: Google Gemini. (2026). AI-generated image created with Google Gemini [Create a logo design. The logo features a stylized emblem in the shape of a strawberry with linear, helical, reverse-strip patterns. The emblem is slanted at a 45-degree angle to the left, and its stem should be red while the base should be green. Below the two lines of the logo, let there be a sans-serif text reading 'Çilek Çikolata' as a logotype consisting of two words.] Google Gemini https://gemini.google.com/

Görsel 10.1: Google Gemini. (2026). AI-generated image created with Google Gemini [Create a packaging design for this uploaded logo design. The packaging should be designed as a rectangular, vertical white chocolate bar wrapper. There should be a vertical red strip on the right and left sides of the packaging. The logo should be placed in the upper center portion of the packaging. The lower middle part of the packaging should say '75% COCOA' in red. Below that, it should say '100% NATURAL' in a smaller red font, and below that, it should say '120 G' in a slightly larger font.] Google Gemini https://gemini.google.com/

Görsel 11: Ideogram.ai. (2026). AI-generated image created with ideogram.ai. [Create a logo design. The logo features a stylized emblem in the shape of a strawberry with linear, helical, reverse-strip patterns. The emblem is slanted at a 45-degree angle to the left, and its stem should be red while the base should be green. Below the two lines of the logo, let there be a sans-serif text reading 'Çilek Çikolata' as a logotype consisting of two words.] Ideogram.ai. https://ideogram.ai/t/

Görsel 11.1: Ideogram.ai. (2026). AI-generated image created with ideogram.ai. [Create a packaging design for this uploaded logo design. The packaging should be designed as a rectangular, vertical white chocolate bar wrapper. There should be a vertical red strip on the right and left sides of the packaging. The logo should be placed in the upper center portion of the packaging. The lower middle part of the packaging should say '75% COCOA' in red. Below that, it should say '100% NATURAL' in a smaller red font, and below that, it should say '120 G' in a slightly larger font.] Ideogram.ai. https://ideogram.ai/t/

Görsel 12: RunwayMl. (2026). AI-generated image created with RunwayMl. [Create a logo design. The logo features a stylized emblem in the shape of a strawberry with linear, helical, reverse-strip patterns. The emblem is slanted at a 45-degree angle to the left, and its stem should be red while the base should be green. Below the two lines of the logo, let there be a sans-serif text reading 'Çilek Çikolata' as a logotype consisting of two words.] RunwayML https://app.runwayml.com/

Görsel 12.1: RunwayMl. (2026). AI-generated image created with RunwayMl. [Create a packaging design for this uploaded logo design. The packaging should be designed as a rectangular, vertical white chocolate bar wrapper. There should be a vertical red strip on the right and left sides of the packaging. The logo should be placed in the upper center portion of the

packaging. The lower middle part of the packaging should say '75% COCOA' in red. Below that, it should say '100% NATURAL' in a smaller red font, and below that, it should say '120 G' in a slightly larger font.] RunwayML https://app.runwayml.com/

Görsel 13: OpenAI. (2026). AI-generated image created with ChatGPT/DALL·E, [Create a logo design. The logo should contain an illustrative emblem designed from an octopus figure. The octopus figure should be navy blue on a white background. The tentacles of the octopus figure should appear in an illustrative style, with curved, asymmetrical balance and order. The octopus figure should be positioned facing the left. The logo should have a round, dark blue ring-shaped outer line. The logotype text should be aligned to the top center section according to the slope of the dark blue circle, and should be written in a curved font with the text 'AQUISO' matching the illustrative style of the emblem. Additionally, separate from the brand name, let there be a sans-serif text consisting of two white words, 'Frozen Seafood', aligned to the bottom center of the navy blue ring.] OpenAI. https://chatgpt.com

Görsel 13.1: OpenAI. (2026). AI-generated image created with ChatGPT/DALL·E, [Create a packaging design for this uploaded logo design. The packaging should be rectangular, vertical, and have two colors. The upper 40% of the packaging should be white, while the lower 60% should be dark blue. Place the logo in the center of the upper white section. Add dark blue 'FROZEN' text below the logo. In the dark blue central part of the packaging, create a white circular background. Create an illustrative composition on a round white background featuring 'a pink-hued octopus extending beyond the edge of the area, a red-hued lobster, a gray fish, a black-hued shelled mussel, and two slices of salmon.' The white text 'SEAFOOD' should appear in the upper part of the white circular background. The text '500g' should appear in the lighter shade of the blue tone on the background in the lower right corner of the packaging.] OpenAI. https://chatgpt.com

Görsel 14: Google Gemini. (2026). AI-generated image created with Google Gemini [Create a logo design. The logo should contain an illustrative emblem designed from an octopus figure. The octopus figure should be navy blue on a white background. The tentacles of the octopus figure should appear in an illustrative style, with curved, asymmetrical balance and order. The octopus figure should be positioned facing the left. The logo should have a round, dark blue ring-shaped outer line. The logotype text should be aligned to the top center section according to the slope of the dark blue circle, and should be written in a curved font with the text 'AQUISO' matching the illustrative style of the emblem. Additionally, separate from the brand name, let there be a sans-serif text consisting of two white words, 'Frozen Seafood', aligned to the bottom center of the navy blue ring.] Google Gemini https://gemini.google.com/

Görsel 14.1: Google Gemini. (2026). AI-generated image created with Google Gemini [Create a packaging design for this uploaded logo design. The packaging should be rectangular, vertical, and have two colors. The upper 40% of the packaging should be white, while the lower 60% should be dark blue. Place the logo in the center of the upper white section. Add dark blue 'FROZEN' text below the logo. In the dark blue central part of the packaging, create a white circular background. Create an illustrative composition on a round white background featuring 'a pink-hued octopus extending beyond the edge of the area, a red-hued lobster, a gray fish, a black-hued shelled mussel, and two slices of salmon.' The white text 'SEAFOOD' should appear in the upper part of the white circular background. The text '500g' should appear in the lighter shade of the blue tone on the background in the lower right corner of the packaging.] Google Gemini https://gemini.google.com/

Görsel 15: Ideogram.ai. (2026). AI-generated image created with ideogram.ai. [Create a logo design. The logo should contain an illustrative emblem designed from an octopus figure. The octopus figure should be navy blue on a white background. The tentacles of the octopus figure should appear in an illustrative style, with curved, asymmetrical balance and order. The octopus figure should be positioned facing the left. The logo should have a round, dark blue ring-shaped outer line. The logotype text should be aligned to the top center section according to the slope of the dark blue circle, and should be written in a curved font with the text 'AQUISO' matching the illustrative style of the emblem. Additionally, separate from the brand name, let there be a sans-serif text consisting of two white words, 'Frozen Seafood', aligned to the bottom center of the navy blue ring.] Ideogram.ai. https://ideogram.ai/t/

Görsel 15.1: Ideogram.ai. (2026). AI-generated image created with ideogram.ai. [Create a packaging design for this uploaded logo design. The packaging should be rectangular, vertical, and have two colors. The upper 40% of the packaging should be white, while the lower 60% should be dark blue. Place the logo in the center of the upper white section. Add dark blue 'FROZEN' text below the logo. In the dark blue central part of the packaging, create a white circular background. Create an illustrative composition on a round white background featuring 'a pink-hued octopus extending beyond the edge of the area, a red-hued lobster, a gray fish, a black-hued shelled mussel, and two slices of salmon.' The white text 'SEAFOOD' should appear in the upper part of the white circular background. The text '500g' should appear in the lighter shade of the blue tone on the background in the lower right corner of the packaging.] Ideogram.ai. https://ideogram.ai/t/

Görsel 16: RunwayMl. (2026). AI-generated image created with RunwayMl. [Create a logo design. The logo should contain an illustrative emblem designed from an octopus figure. The octopus figure should be navy blue on a white background. The tentacles of the octopus figure should appear in an illustrative style, with curved, asymmetrical balance and order. The octopus figure should be positioned facing the left. The logo should have a round, dark blue ring-shaped outer line. The logotype text should be aligned to the top center section according to the slope of the dark blue circle, and should be written in a curved font with the text 'AQUISO' matching the illustrative style of the emblem. Additionally, separate from the brand name, let there be a sans-serif text consisting of two white words, 'Frozen Seafood', aligned to the bottom center of the navy blue ring.] RunwayML https://app.runwayml.com/

Görsel 16.1: RunwayMl. (2026). AI-generated image created with RunwayMl. [Create a packaging design for this uploaded logo design. The packaging should be rectangular, vertical, and have two colors. The upper 40% of the packaging should be white, while the lower 60% should be dark blue. Place the logo in the center of the upper white section. Add dark blue 'FROZEN' text below the logo. In the dark blue central part of the packaging, create a white circular background. Create an illustrative composition on a round white background featuring 'a pink-hued octopus extending beyond the edge of the area, a red-hued lobster, a gray fish, a black-hued shelled mussel, and two slices of salmon.' The white text 'SEAFOOD' should appear in the upper part of the white circular background. The text '500g' should appear in the lighter shade of the blue tone on the background in the lower right corner of the packaging.] RunwayML https://app.runwayml.com/

Görsel 17: OpenAI. (2026). AI-generated image created with ChatGPT/DALL·E, [Create a logo design. The logo should be designed as an orange vertical rectangular shape with only the corners oval. The logo should include an illustrative emblem designed from a fox figure. The emblem should be in the form of an illustrative drawing in white with sharp linear nuances on a vertical orange background. The emblem should be designed with the head and tail of a fox figure. The fox figure's head should be positioned facing vertically to the left. The fox figure's tail should curve from right to left over the top of the head. Inside the orange rectangular shape, in the lower middle part of the emblem, write 'FOXUS' in white, sans-serif font.] OpenAI. https://chatgpt.com

Görsel 17.1: OpenAI. (2026). AI-generated image created with ChatGPT/DALL·E, [Create a packaging design for this uploaded logo design. The packaging should be rectangular, vertical, and in white color. The top part of the packaging should have a locking system, while the bottom part should be square, wide, and have depth. The packaging should be a white-colored filter coffee packaging. The left side of the packaging should be visible and should be orange. Place the logo in the center of the packaging. Add medium-sized orange text reading 'Coffee' below the logo, and place the text '100% Arabica' below it in a small orange square, one line below the other. The lower left corner of the packaging should have a small, orange-colored label reading '500g'.] OpenAI. https://chatgpt.com

Görsel 18: Google Gemini. (2026). AI-generated image created with Google Gemini [Create a logo design. The logo should be designed as an orange vertical rectangular shape with only the corners oval. The logo should include an illustrative emblem designed from a fox figure. The emblem should be in the form of an illustrative drawing in white with sharp linear nuances on a vertical orange background. The emblem should be designed with the head and tail of a fox figure. The fox figure's head should be positioned facing vertically to the left. The fox figure's tail should curve from right to left over the top of the head. Inside the orange rectangular shape, in the lower middle part of the emblem, write 'FOXUS' in white, sans-serif font.] Google Gemini https://gemini.google.com/

Görsel 18.1: Google Gemini. (2026). AI-generated image created with Google Gemini [Create a packaging design for this uploaded logo design. The packaging should be rectangular, vertical, and in white color. The top

part of the packaging should have a locking system, while the bottom part should be square, wide, and have depth. The packaging should be a white-colored filter coffee packaging. The left side of the packaging should be visible and should be orange. Place the logo in the center of the packaging. Add medium-sized orange text reading 'Coffee' below the logo, and place the text '100% Arabica' below it in a small orange square, one line below the other. The lower left corner of the packaging should have a small, orange-colored label reading '500g']. Google Gemini https://gemini.google.com/

Görsel 19: Ideogram.ai. (2026). AI-generated image created with ideogram.ai [Create a logo design. The logo should be designed as an orange vertical rectangular shape with only the corners oval. The logo should include an illustrative emblem designed from a fox figure. The emblem should be in the form of an illustrative drawing in white with sharp linear nuances on a vertical orange background. The emblem should be designed with the head and tail of a fox figure. The fox figure's head should be positioned facing vertically to the left. The fox figure's tail should curve from right to left over the top of the head. Inside the orange rectangular shape, in the lower middle part of the emblem, write 'FOXUS' in white, sans-serif font.] Ideogram.ai. https://ideogram.ai/t/

Görsel 19.1: Ideogram.ai. (2026). AI-generated image created with ideogram.ai [Create a packaging design for this uploaded logo design. The packaging should be rectangular, vertical, and in white color. The top part of the packaging should have a locking system, while the bottom part should be square, wide, and have depth. The packaging should be a white-colored filter coffee packaging. The left side of the packaging should be visible and should be orange. Place the logo in the center of the packaging. Add medium-sized orange text reading 'Coffee' below the logo, and place the text '100% Arabica' below it in a small orange square, one line below the other. The lower left corner of the packaging should have a small, orange-colored label reading '500g'] Ideogram.ai. https://ideogram.ai/t/

Görsel 20: RunwayMl. (2026). AI-generated image created with RunwayMl [Create a logo design. The logo should be designed as an orange vertical rectangular shape with only the corners oval. The logo should include an illustrative emblem designed from a fox figure. The emblem should be in the form of an illustrative drawing in white with sharp linear nuances on a vertical orange background. The emblem should be designed with the head and tail of a fox figure. The fox figure's head should be positioned facing vertically to the left. The fox figure's tail should curve from right to left over the top of the head. Inside the orange rectangular shape, in the lower middle part of the emblem, write 'FOXUS' in white, sans-serif font.] RunwayML https://app.runwayml.com/

Görsel 20.1: RunwayMl. (2026). AI-generated image created with [Create a packaging design for this uploaded logo design. The packaging should be rectangular, vertical, and in white color. The top part of the packaging should have a locking system, while the bottom part should be square, wide, and have depth. The packaging should be a white-colored filter coffee packaging. The left side of the packaging should be visible and should be orange. Place the logo in the center of the packaging. Add medium-sized orange text reading 'Coffee' below the logo, and place the text '100% Arabica' below it in a small orange square, one line below the other. The lower left corner of the packaging should have a small, orange-colored label reading '500g'] RunwayML https://app.runwayml.com/

Türk Dünyası Görsel Kültürünün Dijital Dönüşümü: Ortak Motiflerin Yapay Zekâ ile Yeniden Yorumlanması

Burak Erhan Tarlakazan
Grafik Tasarımı Bölümü (Güzel Sanatlar ve Tasarım Fakültesi)
Kastamonu Üniversitesi
Türkiye
ORCID: 0000-0002-5826-2148

***Özet*— Türk dünyası; tarihsel, dilsel, kültürel ve coğrafi bağlar temelinde birbirleriyle ilişki içinde bulunan ve Türk dili ailesine mensup toplulukların yaşadığı geniş bir kültürel alanı ifade etmektedir. Orta Asya'dan Anadolu'ya, Kafkasya'dan Balkanlar'a ve Sibirya'dan Orta Doğu'ya uzanan bu geniş coğrafyada yaşayan Türk toplulukları; ortak dil kökeni, benzer mitolojik anlatılar, geleneksel sanat anlayışı, motif repertuvarı, yaşam pratikleri ve kolektif hafıza unsurları aracılığıyla birbirine bağlanan çok boyutlu bir kültürel bütünlük oluşturmaktadır. Bu bağlamda Türk dünyası yalnızca bir coğrafi tanımlama değil; aynı zamanda ortak tarihsel deneyimlerin, sembolik anlatım biçimlerinin ve görsel kültür mirasının paylaşıldığı dinamik bir kültürel etkileşim alanı olarak değerlendirilmektedir.**

Bu değerlendirme; ortak motifler üzerinden okunabilen güçlü bir görsel kültür mirasını da kapsamaktadır. Söz konusu ortak motifler, kültürel kimliğin sürekliliğini sağlayan sembolik anlatım biçimleri olarak günümüzde dijital tasarım ve yapay zekâ uygulama ve ortamlarında yeniden yorumlanmaya açık önemli bir üretim alanı potansiyeline sahiptir.

Bu çalışma, Türk dünyasına ait geleneksel görsel kültür mirasının yapay zekâ ile nasıl yeniden üretilebileceğini ve yorumlanabileceğini, özellikle ortak motifler üzerinden ele almayı amaçlamaktadır. Türk dünyasının geniş coğrafyasında şekillenen halı, kilim, dokuma, mimari bezeme, minyatür, süsleme vb. sanatlarında ortaya çıkan zengin bir ortak motif repertuvarına sahiptir. Bu motifler yalnızca estetik birer unsur değil, aynı zamanda kültürel kimlik, kolektif hafıza ve sembolik anlatım biçimlerinin taşıyıcısıdır.

Çalışmada, yapay zekâ ve tasarım ilişkisi bağlamında üretken yapay zekâ araçlarının görsel kültürel mirasın yeniden yorumlanmasındaki potansiyeli incelenmektedir. Özellikle ortak motiflerin yapay zekâ destekli görselleştirme süreçleri aracılığıyla yeniden üretilmesi, çeşitlendirilmesi ve çağdaş grafik tasarım diline aktarılması üzerinde durulmaktadır. Bu doğrultuda, seçilen geleneksel motif örneklerinin dijital ortama aktarılması, veri seti oluşturma süreçleri ve yapay zekâ temelli görsel üretim teknikleri üzerinden deneysel uygulama önerileri geliştirilmektedir.

Araştırma, ortak motiflerin yapay zekâ ile yeniden görselleştirilmesi sürecinin yalnızca teknik bir dönüşüm değil; aynı zamanda kültürel sürekliliğin dijital ortamda yeniden kurulmasına yönelik yaratıcı bir tasarım yaklaşımı sunduğunu ortaya koymaktadır. Bu bağlamda çalışma, Türk dünyası görsel kültürünün dijital çağda sürdürülebilirliğine katkı sağlamayı ve yapay zekâ destekli tasarım süreçlerinin kültürel miras odaklı yeni üretim modelleri geliştirmedeki rolünü tartışmayı hedeflemektedir.

Anahtar Kelimeler—Türk dünyası, görsel kültür, motif, yapay zeka.

(Abstract)— The Turkic world refers to a vast cultural region inhabited by communities belonging to the Turkic language family, which are interconnected through historical, linguistic, cultural, and geographical ties. The Turkic communities living across this vast geography—stretching from Central Asia to Anatolia, the Caucasus to the Balkans, and Siberia to the Middle East—form a multidimensional cultural unity connected through a shared linguistic origin, similar mythological narratives, a traditional understanding of art, a repertoire of motifs, daily practices, and elements of collective memory. In this context, the Turkic world is not merely a geographical designation; it is also regarded as a dynamic space of cultural interaction where shared historical experiences, symbolic forms of expression, and visual cultural heritage are exchanged.

This assessment also encompasses a rich visual cultural heritage that can be interpreted through shared motifs. These shared motifs, as symbolic forms of expression that ensure the continuity of cultural identity, hold significant potential as a field of production open to reinterpretation in today's digital design and artificial intelligence applications and environments.

This study aims to examine how the traditional visual cultural heritage of the Turkic world can be reproduced and interpreted using artificial intelligence, with a particular focus on shared motifs. The vast geography of the Turkic world has given rise to a rich repertoire of shared motifs in arts such as carpets, kilims, textiles, architectural ornamentation, miniatures, and decorative arts. These motifs are not merely aesthetic elements; they also serve as carriers of cultural identity, collective memory, and symbolic forms of expression.

This study examines the potential of generative AI tools in reinterpreting visual cultural heritage within the context of the relationship between artificial intelligence and design. Particular emphasis is placed on the reproduction, diversification, and adaptation of shared motifs into contemporary graphic design language through AI-assisted visualization processes. In this vein, experimental application proposals are developed based on the digitization of selected traditional motif examples, data set creation processes, and AI-based visual production techniques.

The research demonstrates that the process of reimagining shared motifs through artificial intelligence is not merely a technical transformation; it also offers a creative design approach aimed at reestablishing cultural continuity in the digital realm. In this context, the study aims to contribute to the sustainability of the visual culture of the Turkic world in the digital age and to discuss the role of AI-supported design processes in developing new production models focused on cultural heritage.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

(key words)—the Turkic world, visual culture, motifs, artificial intelligence.

I. GİRİŞ

Dijitalleşme, kültürel üretim biçimlerini dönüştüren en önemli dinamiklerden biri hâline gelmiştir. Özellikle yapay zekâ teknolojilerinin gelişimi; sanat, tasarım ve görsel iletişim alanlarında yeni üretim modellerinin ortaya çıkmasına neden olmuş, geleneksel kültürel miras unsurlarının dijital ortamda yeniden yorumlanmasını mümkün kılmıştır. Bu süreç yalnızca teknik bir dönüşüm değil; aynı zamanda kültürel belleğin yeniden temsil edilme biçimlerini etkileyen önemli bir paradigma değişimi olarak değerlendirilmektedir. Günümüzde üretken yapay zekâ sistemleri; görsel veri analizi, desen üretimi, biçim dönüşümü ve stil transferi gibi yöntemlerle geleneksel sanat unsurlarını yeniden üretebilmekte ve çağdaş tasarım pratiklerine entegre edebilmektedir.

Kültürel mirasın dijital ortamda korunması ve yeniden üretilmesi konusu, özellikle görsel kültür araştırmalarında önemli bir çalışma alanı hâline gelmiştir. UNESCO tarafından yayımlanan raporlarda, somut ve somut olmayan kültürel miras unsurlarının dijital teknolojiler aracılığıyla belgelenmesi, korunması ve gelecek kuşaklara aktarılmasının kültürel sürdürülebilirlik açısından önemli olduğu vurgulanmaktadır [1]. Bu bağlamda dijital teknolojiler yalnızca bir arşivleme aracı değil; aynı zamanda kültürel anlatıların yeniden üretildiği yaratıcı platformlar olarak değerlendirilmektedir.

Türk dünyası, ortak tarihsel geçmişi ve kültürel kodları aracılığıyla güçlü bir görsel kültür birikimine sahiptir. Orta Asya'dan Anadolu'ya uzanan geniş coğrafyada gelişen halı, kilim, çini, dokuma, mimari süsleme ve minyatür sanatları; ortak semboller, geometrik düzenler ve motif sistemleri üzerinden kültürel devamlılık sağlamaktadır. Stuart Hall'a göre görsel imgeler, toplumların anlam üretme süreçlerinde kültürel kimliğin kurulmasını sağlayan önemli temsil araçlarıdır [2]. Buna göre geleneksel motifler yalnızca estetik düzenlemeler değil; aynı zamanda kültürel kimliği görünür kılan sembolik anlatım biçimleri olarak değerlendirilmektedir.

Türk kültürüne ait geleneksel motiflerin önemli bir bölümü; hayat ağacı, koçboynuzu, güneş, yıldız, bereket, suyolu ve geometrik örgü sistemleri gibi ortak sembolik yapılardan oluşmaktadır. Bu semboller farklı Türk topluluklarında biçimsel değişiklikler gösterse de ortak kültürel kökenleri yansıtmaktadır. Jan Assmann, kültürel belleği toplumların ortak geçmişini semboller ve ritüeller aracılığıyla geleceğe taşıyan bir aktarım sistemi olarak tanımlamaktadır [3]. Bu açıdan değerlendirildiğinde geleneksel motifler, Türk dünyasının kolektif hafızasını taşıyan görsel kodlar olarak işlev görmektedir.

Son yıllarda Türk dünyasında kültürel mirasın dijitalleştirilmesine yönelik çeşitli projeler geliştirilmiştir. Türk Kültür ve Miras Vakfı tarafından yürütülen dijital katalog çalışmaları kapsamında Türk dünyasına ait kültürel objelerin ve geleneksel sanat unsurlarının dijital ortama aktarılması hedeflenmektedir. Vakıf tarafından yapılan açıklamalarda, ortak kültürel mirasın dijital veri tabanları aracılığıyla korunmasının amaçlandığı belirtilmektedir [4]. Benzer şekilde Kazakistan merkezli Türk Heritage platformu; Türk halklarına ait kültürel miras unsurlarını sanal müzecilik ve dijital arşiv uygulamalarıyla erişilebilir hâle getirmeyi amaçlamaktadır [5]. Bu tür girişimler, geleneksel motiflerin dijital veri setlerine dönüştürülmesi açısından önemli bir altyapı oluşturmaktadır.

Bu noktada üretken yapay zekâ araçları, geleneksel motiflerin yalnızca korunmasını değil; aynı zamanda yeniden yorumlanmasını da mümkün hâle getirmektedir. Yapay zekâ destekli görsel üretim sistemleri; mevcut motif yapılarını analiz ederek yeni kompozisyonlar oluşturabilmekte, farklı kültürel üslupları sentezleyebilmekte ve geleneksel estetik anlayışı çağdaş grafik tasarım diliyle yeniden üretebilmektedir. Böylece kültürel miras, statik bir arşiv nesnesi olmaktan çıkarak dinamik ve yeniden üretilebilir bir görsel tasarım alanına dönüşmektedir.

Bu çalışma, Türk dünyasına ait ortak motiflerin yapay zekâ destekli görsel üretim süreçlerinde nasıl yeniden yorumlanabileceğini incelemektedir. Çalışmada geleneksel motiflerin dijital veri setlerine dönüştürülmesi, üretken yapay zekâ araçlarıyla yeniden görselleştirilmesi ve çağdaş grafik tasarım dili içerisinde yeniden kullanılması üzerine deneysel bir yaklaşım önerilmektedir. Böylece araştırma, kültürel mirasın dijital sürdürülebilirliği ile yapay zekâ destekli tasarım pratikleri arasındaki ilişkiyi tartışmayı amaçlamaktadır.

II. KAVRAMSAL VE KURAMSAL ÇERÇEVE

A. Türk Dünyası Kavramı ve Ortak Kültürel Yapı

Türk Dünyası kavramı; ortak tarihsel köken, dil, kültür, gelenek, inanç sistemi ve kolektif hafıza unsurları temelinde birbirleriyle ilişkili Türk topluluklarının oluşturduğu geniş kültürel coğrafyayı ifade etmektedir. Bu kavram yalnızca etnik veya coğrafi bir tanımlama değil; aynı zamanda tarihsel süreklilik içerisinde şekillenen kültürel aidiyet ve ortak medeniyet anlayışını da kapsamaktadır. Türk dünyası kavramı, Orta Asya'dan Anadolu'ya, Balkanlar'dan Sibirya'ya kadar uzanan geniş coğrafyada yaşayan Türk toplulukları; dil, gelenek, inanç sistemi, sanat anlayışı ve sembolik değerler bakımından önemli ortaklıklar göstermektedir [6].

Kavramsal olarak Türk dünyası; Adriyatik'ten Çin Seddi'ne kadar uzanan bir coğrafyada yaşayan, Türk dillerinin farklı lehçelerini konuşan ve ortak bir tarih bilincine sahip olan halkların sosyo-kültürel ekosistemidir [7]. Bu kavram, 19. yüzyılın sonlarında Gaspıralı İsmail'in "Dilde, fikirde, işte birlik" şiarıyla entelektüel bir zemine oturmuş, Sovyetler Birliği'nin dağılmasının ardından ise jeopolitik ve kurumsal bir gerçekliğe dönüşmüştür. Purtaş (2013) ifade ettiği üzere, 1991 sonrası süreçte bağımsızlığına kavuşan Türk cumhuriyetleri, aralarındaki yapay duvarları yıkarak ortak kimlik arayışına hız vermiş ve bu durum kültürel entegrasyon çalışmalarının merkezini oluşturmuştur [8].

Türk Dünyası içerisinde genel olarak Türkiye, Azerbaycan, Kazakistan, Kırgızistan, Özbekistan ve Türkmenistan gibi bağımsız Türk devletleri yer almakla birlikte; Tatar, Başkurt, Uygur, Gagavuz, Karakalpak ve Ahıska Türkleri gibi farklı topluluklar da bu kültürel bütünün parçaları arasında değerlendirilmektedir. Bu topluluklar arasında görülen ortak kültürel unsurlar; destan geleneği, sözlü anlatım kültürü, halı ve kilim motifleri, mimari süsleme

anlayışı, müzik gelenekleri ve ortak sembolik imgeler üzerinden okunabilmektedir [9].

Dil unsuru, Türk Dünyası'nı bir arada tutan en önemli kültürel bağlardan biri olarak kabul edilmektedir. Türk lehçeleri arasında görülen yapısal benzerlikler, toplulukların düşünce sistemleri ve dünyayı algılama biçimleri arasında da ortaklıklar bulunduğunu göstermektedir. Bu bağlamda dil, yalnızca iletişim aracı değil; kültürel hafızanın, kolektif kimliğin ve toplumsal değerlerin taşıyıcısı olarak değerlendirilmektedir [10]. Özellikle ortak atasözleri, deyimler ve destan anlatıları, Türk toplulukları arasında kültürel sürekliliğin korunmasına katkı sağlamaktadır.

Türk kültüründe ortak semboller ve motifler de kültürel birlikteliğin önemli göstergeleri arasında yer almaktadır. Koçboynuzu, çift başlı kartal, hayat ağacı, güneş motifi ve geometrik yıldız formları gibi semboller; farklı Türk topluluklarının sanat üretimlerinde benzer anlamlarla kullanılmaktadır. Bu semboller çoğunlukla güç, bereket, koruyuculuk, sonsuzluk ve yaşam döngüsü gibi kavramlarla ilişkilendirilmiştir [11]. Özellikle geleneksel dokuma sanatlarında ve mimari süslemelerde görülen geometrik düzen anlayışı, Türk estetik düşüncesinin ortak yönlerinden biri olarak değerlendirilmektedir.

Ortak kültürel yapı, modern dönemde yalnızca geleneksel sanatlar aracılığıyla değil; medya, dijital kültür, tasarım ve kültürel diplomasi faaliyetleriyle de yeniden üretilmektedir. Özellikle Türk Devletleri Teşkilatı gibi kurumlar aracılığıyla yürütülen kültürel iş birlikleri, Türk Dünyası bilincinin güçlenmesine katkı sağlamaktadır. Böylece tarihsel ve kültürel bağlar, çağdaş iletişim araçlarıyla desteklenerek yeni kuşaklara aktarılmaktadır.

B. Görsel Kültür Kavramı

Görsel kültür kavramı, bireylerin çevrelerini görsel imgeler aracılığıyla algılama, yorumlama ve anlamlandırma süreçlerini inceleyen disiplinlerarası bir alan olarak tanımlanmaktadır. Modern toplumlarda iletişimin büyük ölçüde görsel öğeler üzerinden gerçekleşmesi, görsel kültürü yalnızca sanat tarihiyle sınırlı bir inceleme alanı olmaktan çıkararak medya, reklam, sinema, dijital tasarım, mimarlık ve gündelik yaşam pratiklerini kapsayan geniş bir araştırma alanına dönüştürmüştür. Bu bağlamda görsel kültür; imgelerin üretimi, dolaşımı ve tüketimi üzerinden toplumsal anlamların nasıl oluştuğunu incelemektedir [12].

Görsel kültür çalışmaları, görmenin yalnızca biyolojik bir süreç olmadığını; aynı zamanda kültürel, ideolojik ve toplumsal etkilerle şekillenen bir deneyim olduğunu savunmaktadır. Bu yaklaşım doğrultusunda bireylerin imgeleri algılama biçimleri, yaşadıkları toplumun değer yargıları, tarihsel birikimi ve kültürel kodlarıyla doğrudan ilişkilidir. John Berger, görme biçimlerinin kültürel olarak inşa edildiğini belirterek insanların baktıkları şeyleri sahip oldukları bilgi ve inanç sistemleri doğrultusunda yorumladıklarını ifade etmektedir [13]. Böylece imgeler yalnızca estetik nesneler değil; aynı zamanda ideolojik ve kültürel anlam taşıyan göstergeler hâline gelmektedir.

Teknolojik gelişmelerle birlikte görsel kültürün etkisi daha da genişlemiştir. Özellikle televizyon, internet ve sosyal medya platformlarının yaygınlaşmasıyla birlikte bireyler günlük yaşamlarında yoğun bir görsel mesaj akışıyla karşı karşıya kalmaktadır. Bu durum, görsel okuryazarlık kavramını da önemli hâle getirmiştir. Görsel okuryazarlık; bireyin imgeleri çözümleyebilme, eleştirel biçimde değerlendirebilme ve görsel mesajların altında yatan anlamları okuyabilme becerisi olarak tanımlanmaktadır [14]. Günümüzde reklam tasarımları, dijital içerikler ve marka iletişimleri büyük ölçüde görsel göstergeler üzerinden kurgulandığı için görsel kültür çalışmaları grafik tasarım alanıyla da doğrudan ilişki içerisindedir.

Görsel kültür aynı zamanda kimlik oluşumu ve kültürel temsil süreçlerinde de önemli bir role sahiptir. Toplumlar; tarihsel hafızalarını, geleneksel sembollerini ve kültürel değerlerini görsel unsurlar aracılığıyla yeni kuşaklara aktarmaktadır. Özellikle geleneksel motiflerin çağdaş tasarım uygulamalarında yeniden yorumlanması, kültürel kimliğin görsel temsil biçimlerinden biri olarak değerlendirilmektedir. Bu nedenle görsel kültür çalışmaları, yalnızca imgelerin estetik yönünü değil; aynı zamanda toplumsal, politik ve kültürel işlevlerini de incelemektedir [15].

C. Kültürel Motif ve Sembolizm

Kültürel motifler ve semboller, toplumların tarihsel birikimlerini, inanç sistemlerini ve kolektif hafızalarını görsel biçimlerde aktaran önemli kültürel göstergeler arasında yer almaktadır. İnsan toplulukları tarih boyunca duygu, düşünce ve değer sistemlerini çeşitli semboller aracılığıyla ifade etmiş; bu semboller zaman içerisinde kültürel kimliğin taşıyıcı unsurları hâline gelmiştir. Motif kavramı genel anlamıyla belirli bir düşünceyi, inancı ya da estetik anlayışı temsil eden tekrar eden görsel öğeleri ifade ederken; sembolizm bu öğelerin taşıdığı anlam katmanlarını inceleyen bir yaklaşım olarak değerlendirilmektedir [16].

Kültürel semboller yalnızca dekoratif unsurlar değil; aynı zamanda toplumların dünya görüşlerini yansıtan anlam sistemleridir. Carl Gustav Jung, sembollerin insanlığın ortak bilinçdışında yer alan arketiplerle ilişkili olduğunu belirtmekte ve kültürel imgelerin kolektif hafızanın görsel ifadeleri olduğunu savunmaktadır [17]. Bu bağlamda toplumlar tarafından kullanılan hayvan figürleri, geometrik biçimler, bitkisel bezemeler ve mitolojik imgeler; yalnızca estetik tercihler değil, aynı zamanda kültürel anlam taşıyan simgesel anlatımlar olarak değerlendirilmektedir.

Türk kültüründe görülen motif ve semboller de tarihsel süreç içerisinde şekillenmiş önemli anlam katmanlarına sahiptir. Özellikle koçboynuzu, hayat ağacı, çift başlı kartal, güneş motifi ve geometrik yıldız düzenlemeleri; güç, bereket, koruma, sonsuzluk ve yaşam döngüsü gibi kavramlarla ilişkilendirilmektedir. Bu semboller geleneksel halı-kilim sanatında, mimari süslemelerde, mezar taşlarında ve el sanatlarında yaygın biçimde kullanılmıştır [18]. Örneğin koçboynuzu motifi çoğunlukla güç ve kahramanlık sembolü olarak yorumlanırken, hayat ağacı motifi evrenin sürekliliğini ve yaşamın döngüsel yapısını temsil etmektedir.

Motiflerin kültürel aktarım sürecindeki rolü, görsel iletişim açısından da önem taşımaktadır. Kültürel motifler, geçmişten gelen estetik anlayışların çağdaş tasarım pratikleri içerisinde yeniden yorumlanmasına olanak sağlamaktadır. Özellikle grafik tasarım, ambalaj tasarımı ve dijital medya alanlarında geleneksel motiflerin modern görsel dille yeniden üretilmesi, kültürel kimliğin korunması ve görünür kılınması açısından önemli bir yaklaşım olarak değerlendirilmektedir. Göstergebilim perspektifine göre semboller, kültür içerisinde ortak anlam üretimini sağlayan göstergeler sistemi olarak işlev görmektedir [19]. Bu nedenle kültürel motifler yalnızca

estetik öğeler değil; aynı zamanda toplumsal hafızayı taşıyan görsel anlatım araçlarıdır.

Günümüzde küreselleşme ve dijitalleşmenin etkisiyle kültürel semboller farklı bağlamlarda yeniden üretilmekte ve uluslararası görsel iletişim içerisinde daha görünür hâle gelmektedir. Ancak bu süreç, bazı geleneksel motiflerin özgün anlamlarının dönüşmesine ya da ticarileşmesine de neden olabilmektedir. Bu nedenle kültürel sembollerin çağdaş tasarım uygulamalarında kullanımı sırasında tarihsel ve kültürel bağlamlarının korunması önem taşımaktadır.

D. Dijitalleşme ve Kültürel Üretim İlişkisi

Teknolojinin tarihsel süreç içerisindeki gelişimiyle birlikte toplumların yaşam biçimleri ve üretim pratikleri köklü bir değişim sürecine girmiştir. Geleneksel kültür anlayışı, bilgi ve iletişim teknolojilerindeki hızlı ilerlemeler ve internetin yaşamın her alanına nüfuz etmesiyle yerini büyük ölçüde "dijital kültür" kavramına bırakmıştır [20]. En temel anlamıyla dijitalleşme; fiziksel çabanın en aza indirilip verimliliğin en üst düzeye çıkarılması amacıyla var olan kaynakların ve süreçlerin dijital sistemlerle (bilgisayar dilleri ve kodlarla) yeniden üretilmesi ve ikame edilmesi sürecidir [21]. Bu dönüşüm, kültürel üretimin sınırlarını, aktarım yollarını ve anlam yaratma mekanizmalarını yeniden tanımlamaktadır.

Dijitalleşme ekseninde kültürel üretim, yaratıcı ekonomi (creative economy) ve kültürel ve yaratıcı endüstriler çerçevesinde yeni bir yapıya bürünmüştür. Geleneksel sanayi ekonomisinde toprak, emek ve sermaye temel üretim girdileri iken; günümüz yaratıcı ekonomisinde kültürel endüstriler, doğrudan fikirler dünyasından anlamlar ve değerler üretmektedir. Bu bağlamda, bireysel yaratıcılığa dayalı kültürel üretim, yalnızca izole bir sanatsal faaliyet olmaktan çıkarak, dijital ekosistemler içerisine yerleşmiş ve kurumsal bir nitelik kazanmıştır. Kültürel ürünlerin yaratımı, dağıtımı ve ticarileşmesi aşamaları, artık büyük ölçüde teknolojik altyapılar ve dijital platformlar üzerinden yürütülmektedir [22].

Dijital çağ, yalnızca yeni kültürel eserlerin üretimine değil, var olan kültürel mirasın yeniden yorumlanmasına da olanak tanımaktadır. Sanal gerçeklik (VR), artırılmış gerçeklik (AR), üç boyutlu modelleme ve dijital arşivler gibi teknolojik araçlar, kültür varlıklarının zaman ve mekân kısıtlamalarından bağımsız olarak çok daha geniş kitlelere ulaştırılmasını sağlamaktadır [23]. Bu sayede kültürel üretim süreci; geçmişin kültürel varlıklarının dijital oyunlarda, sanal turlarda ve zenginleştirilmiş müze deneyimlerinde (örneğin; sanal müzelerdeki etkileşimli kurulumlar) kullanıcıyla buluşturularak yeniden üretilmesini de kapsar hale gelmiştir [24]. Dijitalleşme ile kültürel üretim ilişkisindeki en dramatik değişimlerden bir diğeri, geleneksel tüketici ve üretici rollerinin iç içe geçmesidir. Medya katılımının artmasıyla birlikte, çevrim içi kültür, bilgisayarlar ve insanlar arasındaki etkileşimlerin anlamlı sosyal ilişkilere dönüşmesine zemin hazırlamış; kullanıcıların doğrudan kendi içeriklerini yarattıkları interaktif bir yapı ortaya çıkmıştır [25]. Bireyler artık bloglar, podcastler ve çeşitli dijital sosyal iletişim araçları vasıtasıyla kendi kültürel içeriklerini oluşturabilmekte ve küresel ölçekte paylaşıma sunabilmektedir. Ancak, yüksek katılımlı bu yeni kültürel üretim biçimi birtakım riskleri de beraberinde getirmektedir. Dijital mecraların anonim ve denetimsiz yapısı, sanal dünyanın kendine has popülerlik arayışı ve yüzeysel tüketim alışkanlıkları; kültürel üretimin bir metaya dönüşmesine, kültürel ve ahlaki değer algısının ticari bir boyuta kayarak yozlaşmasına zemin hazırlayabilmektedir [26]. Sonuç olarak, dijitalleşme ve kültürel üretim ilişkisi, kültürün pasif bir şekilde teknolojiye aktarılmasından ibaret değildir. Bu ilişki; kültürün dijital araçlarla yeniden üretildiği, bireylerin aktif içerik üreticisi konumuna yükseldiği ve yaratıcı endüstriler aracılığıyla ekonomik ve sosyokültürel değerlerin eş zamanlı olarak inşa edildiği çok boyutlu bir etkileşimdir. Kültürel üretimin dijital teknolojiler aracılığıyla geçirdiği bu yapısal dönüşüm, bilhassa otonom sistemlerin ve yapay zekâ teknolojilerinin yaratım süreçlerine dâhil olmasıyla birlikte kavramsal olarak yeni tartışma alanları açmaktadır.

E. Yapay Zekâ ve Sanat

Yapay zekânın (YZ) sanatsal üretime dâhil olması; sanatın tanımı, yazarın niyeti ve yaratıcılık gibi estetiğin temel kavramlarını kökünden sarsan felsefi ve akademik bir tartışma yaratmıştır [27]. Bu alandaki literatürün merkezinde yer alan en belirgin kırılma noktalarından biri, eser üzerindeki hak sahipliği ve yazarın kimliği sorunudur [28]. Örneğin, 2018 yılında Christie's müzayedesinde 432.500 dolara satılan ve algoritmik bir süreçle üretilen Portrait of Edmond de Belamy adlı eser; yaratıcının kodu yazan kişi mi, veri setini belirleyen insan mı yoksa algoritmanın kendisi mi olduğuna dair devasa bir yasal ve kavramsal belirsizliği gün yüzüne çıkarmıştır [29].

Geleneksel estetik anlayışına göre sanat eseri, insanın duygusal derinliğini, bilincini ve kişisel niyetini barındırmak zorundadır [30]. Yapay zekâ, öznel bir yaşam deneyiminden, duygusal zekâdan ve bilinçten yoksun mekanik bir sistem olduğu için, birçok eleştirmen onun ürettiği eserlerin otantik sanattan ziyade Jean Baudrillard'ın "simülakr" kavramıyla örtüşen ruhsuz kopyalar olduğunu savunmaktadır [31], [32].

Muhafazakâr estetik eleştirmeni Roger Scruton gibi isimler, dijital platformların ve algoritmaların sunduğu kısayolların, geleneksel sanatın gerektirdiği katı disiplin ve zanaatkârlığı zedeleyerek şablonlaşmış eserler ortaya çıkardığını ileri sürmektedir [33]. Bu sert eleştirilerin karşısında konumlanan Bilişsel Yaratıcılık alanı ise, makinelerin tarafsız gözlemciler tarafından "yaratıcı" olarak nitelendirilebilecek karmaşık estetik kararlar alabildiğini ve yepyeni kavramsal ufuklar açabildiğini iddia etmektedir [34]. Bu yaklaşım, sanatın yalnızca bireysel ve organik bir eylem olmaktan çıkıp, insan ve makinenin birleştiği kolektif bir olguya (Co-Cre-AI-tion) dönüştüğünü vurgular [35], [36].

İnşa edilen bu yeni ontolojik düzlemde yapay zekâ, insanın organik sınırlarını aşmasını sağlayan posthümanist (insan ötesi) veya "artırılmış insan" (augmented human) modelinin koparılamaz bir parçasıdır [37]. Söz konusu melez iş birliği sayesinde yapay zekâ; sadece bilindik stilleri taklit eden bir araç değil, insan sezgisi ve muhakemesiyle birleşerek sanatsal ufukları ve ihtimalleri genişleten yenilikçi bir partner olarak kabul edilmektedir [38].

Yapay zekâ ve sanat kesişimi, güzellik algısı ve estetik yargı süreçlerinin baştan aşağı yeniden yapılandırılmasını zorunlu kılmaktadır [39]. Geleneksel olarak sanat tarihçisi veya eleştirmenin öznel içgüdüsüne bırakılan estetik değerlendirmeler, bilgisayarlı görme (computer vision) ve nicel veri analizi gibi yöntemlerin devreye girmesiyle "hesaplanabilir güzellik" (computational beauty) kavramını doğurmuştur [40]. Felsefi bir perspektiften bakıldığında; yapay zekânın tasarladığı algoritmaların karmaşıklığı, uyumu ve problem çözme kapasitesi, başlı başına yeni ve çağdaş bir

güzellik formu olarak literatürde kendine yer bulmaya başlamış ve sanatın ontolojik sınırlarını geleneksel tanımların çok ötesine taşımıştır [41].

III. TÜRK DÜNYASINDA ORTAK GÖRSEL MOTİFLER

Türk dünyasında inanç sistemleri, mitoloji ve sosyo-kültürel yaşam, sanata çok çeşitli figürler ve motifler aracılığıyla yansımıştır. Göçebe kültürden yerleşik hayata geçişle birlikte; dokumalardan çinilere, mezar taşlarından mimari eserlere kadar pek çok alanda varlığını sürdüren ortak semboller, derin felsefi anlamlar ve kültürel kodlar taşımaktadır. Bu çalışma, belirlenen kültürel kodları ve temel motifleri alan yazın çerçevesinde akademik bir perspektifle tartışmaktadır.

Koçboynuzu ve Eli Belinde Sembolleri Türk dokuma sanatında ve maddi kültür ürünlerinde köklü bir geçmişe sahip olan koçboynuzu motifi, antik çağlardan itibaren gücün, bereketin, kahramanlığın ve eril enerjinin (erkekliğin) simgesi olarak kabul edilmektedir. Tarihsel süreçte insanoğlunun korunma amacıyla ev kapılarına astığı veya bir tılsım olarak üzerinde taşıdığı koçboynuzu, zamanla stilize edilerek dokuma tezgâhlarında, halılarda ve mezar taşlarında yaşamın bir parçası haline gelmiştir [42].

Bu eril gücü dengede tutan ve tamamlayan "eli belinde" motifi ise analığın, dişiliğin ve doğurganlığın evrensel sembolüdür. Koçboynuzu ve eli belinde motiflerinin geleneksel sanat eserlerinde sıklıkla bir arada kullanılması; evrendeki dişil ve eril zıtlığın birliğini, aile mutluluğunu ve sonsuz bereketi ifade etmektedir [43].

Hayat Ağacı ve Çift Başlı Kartal Sembolleri Eski Türk inançlarından itibaren dünyanın merkezinde bulunduğuna inanılan hayat ağacı; kökleriyle yeraltını, gövdesiyle yeryüzünü ve dallarıyla gökyüzünü birbirine bağlayan kozmik bir eksendir [44]. Türk mitolojisinde doğurganlık, yaşam enerjisi ve türeyişin kaynağı olan hayat ağacı, çoğunlukla Tanrıça Umay inancıyla da yakından ilişkilendirilmiştir [45], [46].

Zamanla İslamiyet'in kabulüyle birlikte Anadolu'da stilize edilmiş hurma ve servi ağaçları, hayat ağacı formunda kullanılmıştır. Servi ağacı yeşil kalmasıyla ebediyeti ve vahdeti simgelerken, hurma ağacı ise cenneti ve bereketi temsil etmiştir [47], [48].

Hayat ağacı figürleri, Erzurum Çifte Minareli Medrese ve Yakutiye Medresesi gibi anıtsal Selçuklu mimarisinde son derece ikonografik kompozisyonlarla işlenmiştir. Bu tasvirlerde hayat ağacı; dallarında cennet kuşları, alt kısmında koruyucu ruhları simgeleyen ejderhalar ve tepesinde hakimiyeti, gücü ve göksel egemenliği temsil eden çift başlı kartal ile birlikte resmedilmiştir [49].

KünAy (Güneş ve Yıldız), Geometrik Geçmeler ve Yıldız Formları (Girih / Selçuklu Yıldızı) Göksel unsurların başında gelen güneş ve yıldız (KünAy) formları, Türk topluluklarında yön gösterici bir rehber ve kutsiyet atfedilen kozmik işaretler olmuştur [50]. Seramikler ve metal işçiliklerinde güneş, aydınlığın ve karanlığa karşı zaferin sembolü olarak tasvir edilmiş; aynı zamanda kartal, göksel kuşlar ve aslan gibi figürlerle eşleştirilerek yeryüzündeki kudreti vurgulamıştır [51].

Yıldız formları arasında bilhassa iki karenin çapraz şekilde üst üste geçmesiyle elde edilen sekiz köşeli yıldız (Selçuklu Yıldızı) en belirgin sembollerden biridir. Bu yıldız formu, Tanrı'nın dört ana ve dört ara yön olmak üzere tüm evrene (sekiz bucağa) olan mutlak hâkimiyetini temsil etmektedir [52], [53].

Öte yandan, Ortaçağ İslam ve Selçuklu döneminde günümüzde genel bir terim olan "geometrik geçmeler" veya "girih" adıyla anılan örgülü yapılar, dönemin yazılı kaynaklarında "mushabak" veya "shabaka" (ağ, kafes) gibi terimlerle adlandırılmıştır. Sekiz köşeli yıldız (Peygamberlerin / Süleyman'ın Mührü) biçimindeki gözeneklerle bezenen bu ağ sistemleri, saray duvarlarında yer aldığında eseri yaptıran yöneticinin "Adil Hükümdar" (İkinci Süleyman) olduğuna gönderme yapmıştır. Tasarımları birbirine bağlayan kesintisiz geçmeler ve düğümler (al-i'tiqad) ise inananların dini inançlara, toplumsal birliğe ve yaratıcıya olan kopmaz itikadını sembolize etmiştir [54].

Kurt (Bozkurt) Sembolleri Türk kültüründe, mitolojisinde ve kolektif hafızasında ana figürlerden olan kurt (bozkurt), el sanatlarında ve Anadolu dokumalarında "kurt ağzı" veya "kurt izi" şeklinde isimlendirilen motiflerle yaşatılmıştır. Kurdun doğada hayatta kalabilmesi ve gece karanlığında görebilme yetisi, bu motife ışık, güneş ve iyimserlik anlamları kazandırmıştır. Çetin tabiat şartlarına ve yaban hayatına karşı güvenlik sağladığına inanılan, kötülüklerden korunma amaçlı dokunan kurt ağzı motifi; aynı zamanda kahramanlığın, cesaretin, sadakatin, gücün ve doğruluğun sembolü olarak kabul görmüştür [55].

Türk kültür tarihi boyunca çeşitli malzeme ve yüzeylerde kullanılan bu zengin ikonografik semboller, toplumun dünya görüşünü, inanç kodlarını, doğaüstü ve kozmik varlıklarla kurduğu felsefi bağları yansıtan temel iletişim araçları olmuştur.

IV. YAPAY ZEKÂ VE GÖRSEL ÜRETİM TEKNOLOJİLERİ

A. Generatif yapay zekâ (GAN, diffusion modelleri vb.)

Yüksek boyutlu ve karmaşık gerçek dünya verilerini modelleme potansiyelleri nedeniyle üretken yapay zeka modelleri, bilgisayarlı görü ve görüntü işleme alanlarında son yıllarda büyük bir akademik ilgi odağı haline gelmiştir [56]. Bu alandaki en büyük kırılmalardan biri, 2014 yılında Ian Goodfellow ve arkadaşları tarafından Üretken Çekişmeli Ağların (Generative Adversarial Networks - GAN) literatüre sunulmasıyla yaşanmıştır [57], [58].

Temelinde oyun teorisi prensibi yatan GAN'lar; veri dağılımını yakalamayı hedefleyen bir üretici (generator) ve üretilen sentetik verileri gerçek verilerden ayırt etmeye çalışan bir ayrıştırıcı (discriminator) olmak üzere iki sinir ağının, sıfır toplamlı (zero-sum) bir minimax oyunu çerçevesinde eşzamanlı olarak eğitilmesine dayanmaktadır [59], [60].

Klasik GAN mimarisi devrim niteliğinde olsa da; eğitim sürecindeki istikrarsızlıklar, modelin doğru biçimde yakınsamaması (non-convergence) ve üreticinin veri çeşitliliğini yansıtmayıp sürekli aynı veya benzer çıktıları ürettiği "mod çökmesi" (mode collapse) gibi kronik teorik problemlerle karşılaşmıştır [61], [62]. Literatür, bu sorunları aşmak ve daha yüksek kalitede görüntüler elde etmek için çeşitli mimari iyileştirmelere tanık olmuştur [63]. Görüntü sentezleme süreçlerini daha istikrarlı hale getirmek için tam bağlantılı (fully-connected) katmanlar yerine derin evrişimli sinir ağlarını (CNN) entegre eden DCGAN (Derin Evrişimli GAN) tanıtılmıştır [64], [65]. Ağırlık kırpma (weight

clipping) tekniği yerine gradiyen cezası (gradient penalty) ve Wasserstein mesafesi formülasyonunu kullanan WGAN ve WGAN-GP modelleri ise, eğitim kararlılığını artırarak mod çökmesi sorununa matematiksel olarak daha sağlam çözümler getirmiştir [66], [67]. Görsel çözünürlüğünü ve üretim kalitesini artırmak adına geliştirilen ProGAN, üretici ve ayrıştırıcının katmanlarını eğitim süresince aşamalı (progressive) olarak büyüterek donanım optimizasyonu eşliğinde yüksek çözünürlüklü görüntülerin dengeli bir şekilde sentezlenmesini sağlamıştır [68], [69].

ProGAN'ın temelleri üzerine inşa edilen StyleGAN ise, üst düzey özniteliklerin (örneğin poz veya kimlik) stokastik varyasyonlardan mimari olarak ayrılmasını sağlayarak, sentezlenen görüntüler üzerinde benzeri görülmemiş bir stil kontrol mekanizması sunmuş ve fotogerçekçi sentezde sınırları yeniden belirlemiştir [70], [71]. GAN modellerinin başlangıçta sunduğu koşulsuz ve rastgele veri üretimini kontrol edilebilir hale getirmek amacıyla Koşullu GAN (Conditional GAN - cGAN) mimarisi geliştirilmiş, bu sayede sınıflar veya belirli etiketler gibi ek bilgiler algoritmaya koşullanarak üretim süreci özel amaçlara yönlendirilmiştir [72], [73]. cGAN temelinde yükselen Pix2Pix modeli, eşleştirilmiş (paired) veri setlerini kullanarak eskizden fotoğrafa, haritadan uydu görüntüsüne veya gece-gündüz dönüşümüne olanak tanıyan alandan alana (image-to-image) doğrudan çeviri görevlerinde büyük bir başarı yakalamıştır [74], [75]. Eğitim süreci için mutlak eşleştirilmiş piksel veri seti bulmanın getirdiği pratik zorlukları ortadan kaldıran CycleGAN modeli ise "döngüsel tutarlılık kaybı" (cycle consistency loss) mekanizmasıyla eşleştirilmemiş veriler (örneğin doğa fotoğraflarından mevsimsel değişimler veya atlardan zebralara çeviri) üzerinde son derece başarılı otonom çeviriler yapabilmiştir [76], [77].

Güncel literatürde GAN'ların mimari üstünlüğüne karşı en güçlü yaklaşım, Denoising Diffusion Probabilistic Models (DDPM) gibi difüzyon tabanlı olasılıksal modellerden gelmiştir; bu modeller Markov zinciri formülasyonu yoluyla gürültüden arındırma (denoising) işlemi uygulayarak mod çökmesi handikabını aşmış ve son derece yüksek görsel doğruluk (fidelity) sunmuştur [78], [79]. Ancak difüzyon modellerinin doğrudan piksel uzayında değerlendirilmesinin oluşturduğu muazzam hesaplama ve zaman maliyeti, araştırmacıları bu süreci önceden eğitilmiş bir otomatik kodlayıcının (autoencoder) algısal olarak eşdeğer ancak veri boyutu açısından çok daha düşük olan "gizil uzayında" (latent space) gerçekleştirmeye itmiştir. Bu doğrultuda geliştirilen Gizil Difüzyon Modelleri (Latent Diffusion Models - LDM), bilgisayımsal donanım maliyetlerini belirgin biçimde azaltırken ayrıntı koruma ve sentezleme kalitesini de en üst düzeye çıkarmıştır [80]. Metinden görüntüye (text-to-image) sentezleme aşamasında ise CLIP (Contrastive Language-Image Pre-training) modeli, doğal dil denetimi üzerinden çok miktarda metin ve görüntüyü aynı semantik uzayda hizalayarak yeni bir devrim yaratmıştır [81]. CLIP modelinin bu güçlü gizil temsilini otonom difüzyon kod çözücüleriyle birleştiren unCLIP (DALL-E 2) gibi hiyerarşik sistemler, dille yönlendirilen görsel varyasyonlarda, yüksek çeşitlilik ve fotogerçekçilikle birleşen üst düzey bir metin-kavramsal esneklik sağlamıştır [82]. Bu ileri düzey üretken yapay zeka mimarileri yalnızca dijital sanat veya kreatif tasarım (örneğin Stable Diffusion ve NovelAI) odaklı değil [83] aynı zamanda sentetik veri çoğaltımından medikal işlemlere kadar birçok akademik ve endüstriyel disiplinde standart haline gelmiştir [84].

B. Sanatsal Üretimde Yapay Zekâ Kullanımı

Yapay zekâ (YZ) ve büyük dil modelleri, günümüzde sanatsal üretim pratiklerinde salt mekanik bir araç olmaktan çıkarak yaratıcı sürecin aktif bir ortağı ve iş birlikçisi hâline gelmiştir [85]. DALL-E, Midjourney, Wombo Dream, Prisma ve ChatGPT gibi üretken yapay zekâ (generative AI) platformları; tasarımcılara ve sanatçılara fikir yürütme, beyin fırtınası yapma ve farklı stilleri sentezleme aşamalarında eşsiz olanaklar sunmaktadır [86], [87].

Bu araçlar, karmaşık metin tabanlı komutları (prompt) saniyeler içinde görsel kompozisyonlara çevirerek yaratım sürecini hızlandırmakta ve sanatçıların geleneksel becerilerinin sınırlarını aşarak soyut tasarımlar yapabilmelerine olanak tanımaktadır [88] , [89].

Görsel sanatlarda yapay zekâ kullanımı, özellikle Üretken Çekişmeli Ağlar (GAN) ve Yaratıcı Çekişmeli Ağlar (CAN) gibi derin öğrenme (deep learning) algoritmalarıyla tamamen yeni bir estetik boyut kazanmıştır. Sanatçılar, devasa sanat tarihi veri setlerini algoritmalarla analiz ederek daha önce görülmemiş melez ve yenilikçi sanat formları inşa etmektedirler [90]. Örneğin, 14. ve 20. yüzyıllar arasındaki 15.000 portrenin analiziyle GAN teknolojisi kullanılarak üretilen ve Christie's müzayedesinde satılan Portrait of Edmond de Belamy adlı tablo ya da Rembrandt'ın 300'den fazla eserinin taranıp karakteristik özelliklerinin öğrenilmesiyle ortaya çıkarılan The Next Rembrandt projesi, bu mekanize yaratıcılığın en çarpıcı örneklerindendir [91], [92]. Uygulama alanları yalnızca resimle sınırlı kalmamaktadır; S. Eaton ve H.B. Snell gibi sanatçılar iki boyutlu verileri yapay zekâ ile üç boyutlu heykellere dönüştürürken, AIVA gibi programlar senfonik müzik bestelemede kullanılmaktadır [93]. Ayrıca yapay zekâ, edebiyat alanında otomatik şiir üretimiyle felsefi ve estetik değerleri formüle edebilmekte [94] ve bilgisayarlı görme (computer vision) ile klasik sanat eserlerinin üslup, renk ve fırça darbesi analizlerini yaparak sanatı sayısal olarak ölçülebilir bir yapıya kavuşturmaktadır [95].

Sanatsal üretimde yapay zekânın artan bu rolü, "yaratıcı özne" kavramını da yapısal olarak dönüştürmektedir. Teknik, fiziksel ve efor gerektiren süreçlerin büyük ölçüde makinelere devredilmesiyle birlikte, sanatçının asıl sermayesi fiziksel beceriden ziyade "fikir", "niyet" ve "kavram" tasarımı hâline gelmiştir [96]. Bu durum, insan ve makinenin simbiyotik bir ilişki kurduğu, "artırılmış insan" (augmented human) veya posthümanist bir sanat ontolojisinin doğmasına yol açmıştır [97].

Tasarım stüdyolarında yapılan araştırmalar, yapay zekânın ilham ve kompozisyon üretmede oldukça başarılı olduğunu; ancak sanatın merkezinde yer alan organik özgünlük, duygusal derinlik, kültürel bağlam ve niyet gibi insani unsurların yerini dolduramadığını vurgulamaktadır [98], [99].

Yapay zekâ mekanik bir üretim hızı ve sonsuz varyasyon sağlasa da, nihayetinde sanatçı bir küratör veya kolaylaştırıcı (facilitator) rolünü üstlenerek bu çıktıları insani bir süzgeçten geçirmekte ve eserine dönüştürmektedir [100]. Tüm bu yenilikçi potansiyeline rağmen, yapay zekânın sanatsal üretimde araçsallaştırılması etik, kültürel ve hukuki kaygıları da beraberinde getirmektedir. Telif hakkı belirsizlikleri, algoritmik önyargıların (bias) yeniden üretimi ve "deepfake"

gibi manipülatif sentetik imajların kullanımı başlıca etik sorunlar arasındadır [101], [102]. Eğitim ve pratik alanında ise öğrencilerin yapay zekâ destekli üretimlere aşırı bağımlı hâle gelmesinin, sanatsal asimilasyona, şablonlaşmış stillere ve özgünlük kaybına (degradation) yol açabileceği uyarısı yapılmaktadır [103]. Öte yandan bazı çağdaş sanatçılar, büyük veri gözetimini (surveillance) ve makine öğreniminin sosyal etkilerini eleştirmek amacıyla bizzat yapay zekâyı kullanarak izleyicide bir tür "yabancılaştırma" (defamiliarization) yaratmakta ve dijital verinin etik sınırlarını sanat pratiği üzerinden sorgulatmaktadırlar [104].

C. Kültürel Verinin Dijitalleşmesi

Kültürel verinin dijitalleşmesi, somut ve somut olmayan miras unsurlarının modern teknolojilerle korunmasını, yeniden yorumlanmasını ve gelecek nesillere aktarılmasını sağlayan çok boyutlu bir süreçtir [105].

Bu süreç; verilerin dijital ortama aktarılmasından yapay zekâ için işlenmesine, mevcut girişimlerden telif ve temsil sorunlarına kadar geniş bir akademik tartışma alanını kapsamaktadır.

Geleneksel kültürel motiflerin ve nesnelerin bilgisayar ortamına aktarılmasında birçok teknolojik araç kullanılmaktadır. Mimari yapılar ve üç boyutlu nesneler için lazer tarama, 3 boyutlu (3B) nokta bulutları, karasal ve hava fotogrametrisi (dijital fotogrametri) gibi yöntemler ön plana çıkmaktadır [106], [107].

Geleneksel sanatlara ait desen ve motiflerin dijitalleştirilmesinde ise görüntülerin otomatik segmentasyonu (bölümlenmesi), nesne hedefli tarama ve desen modelleme gibi bilgisayar destekli vektörel ve geometrik soyutlama tekniklerine başvurulmaktadır [108].

Geleneksel verilerin makine öğrenmesi algoritmaları tarafından işlenebilmesi için yapılandırılmış veri kümelerine dönüştürülmesi gerekmektedir. Makinece okunabilir koleksiyonlar yaratmak amacıyla verilerin öncelikle optik karakter tanıma (OCR) yazılımlarıyla metinleştirilmesi ve XML, RDF, CSV gibi açık formatlarda düzenlenmesi şarttır [109], [110]. Görüntü ve metin tabanlı kültürel veriler toplanıp temizlendikten sonra, yapay zekâ eğitimine uygun olarak duygu, tür veya konsept açısından etiketlenir (semantik etiketleme) ve sınıflandırma modelleri oluşturulur [111], [112]. Dijitalleşme alanında gerek küresel gerekse ulusal düzeyde pek çok girişim mevcuttur. Küresel ölçekte UNESCO'nun kültürel mirasın dijitalleştirilmesine yönelik yönergeleri standartları belirlerken [113] Atina'daki Akropolis Müzesi, Roma'daki Ara Pacis ve Türkiye'deki Topkapı Halı Müzesi gibi müzeler sanal ve artırılmış gerçeklik (AR/VR) uygulamaları ve dijital arşiv projeleri yürütmektedir [114].

Üniversite projeleri bağlamında dijital insani bilimler laboratuvarları (örneğin Sabancı Üniversitesi DH Lab) çeşitli analizler gerçekleştirirken [115] OpenHeritage3D ve CyArk gibi uluslararası oluşumlar, 3B kültürel miras veri setlerini açık erişimli platformlarda küresel paylaşıma sunmaktadır [116]. Veri Kümelerinin Coğrafi ve Kültürel Dağılımı Kültürel mirasa ait dijital veri kümelerinde ve yapay zekâ eğitim verilerinde ciddi temsil sorunları bulunmaktadır. Örneğin OpenHeritage3D platformundaki dijital mirasların büyük bir kısmı Asya ve Kuzey Amerika kıtalarında yoğunlaşmakta; diğer coğrafyalar veri setlerinde eşitsiz bir dağılımla, azınlıkta kalmaktadır [117].

Veri kaynaklarındaki bu eksiklikler ve önyargılar, yapay zekâ algoritmalarının kültürel miras nesnelerini yanlış yorumlamasına ve kültürel bağlamda zararlı stereotiplerin pekiştirilmesine neden olabilmektedir [118]. Küresel belleğin inşası için tüm dillerin, ulusların ve kültürlerin dijital ortamda adil bir şekilde temsil edilmesi zaruridir [119]. Dijitalleşme süreci, fikri mülkiyet ve telif hakları ile ilgili önemli yasal engeller barındırmaktadır [120]. Kültürel miras kurumları eserleri dijitalleştirirken, telif hakkı ile korunan materyallerin sahiplerinden yasal izinler almakta veya anonim/sahipsiz olan "öksüz eserlere" yönelik yazılı idari politikalar geliştirmek zorundadır [121]. Ayrıca piyasanın dijital mecralara hakim olmasıyla kültürün metalaşması ve geleneksel değerlerin ticarileşme riskine karşı kültürel mülkiyetin güvenliğinin sağlanması akademik bir tartışma konusudur [122].

V. YÖNTEM

Bu çalışma, Türk dünyası ortak görsel kültürünün dijital dönüşümünü incelemeyi amaçlayan nitel bir araştırma olup, yöntem olarak örnek olay (case study) incelemesi ve görsel göstergebilimsel/içerik analizi desenlerini benimsemektedir.

Araştırmanın çalışma evrenini Türk dünyası coğrafyası; örneklemini ise kurumsal ve siyasi düzeyde en üst temsiliyete sahip olan, Türk Devletleri Teşkilatı (TDT) kurucu ve tam üye devletleri oluşturmaktadır. Bu doğrultuda; Türkiye, Azerbaycan, Kazakistan, Özbekistan, Kırgızistan ve Türkmenistan olmak üzere altı bağımsız Türk devleti araştırmanın coğrafi ve kültürel sınırları olarak belirlenmiştir. Bu ülkelerin seçilmesindeki temel gerekçe, hem Oğuz hem Kıpçak/Karluk boylarının geleneksel sanat miraslarını temsil etmeleri ve ortak bir sembolik dil havuzuna sahip olmalarıdır.

A. Veri Toplama Süreci ve Motif Belirleme (Taksonomi)

Araştırma kapsamında incelenecek motiflerin belirlenmesinde 'amaçlı örnekleme' (purposive sampling) yöntemi kullanılmıştır. Seçilen motiflerin analize dâhil edilebilmesi için şu üç temel kriteri karşılaması şartı aranmıştır:

1. Söz konusu altı ülkenin en az dördünün geleneksel el sanatlarında (halı, kilim, keçe, mimari, tekstil vb.) aktif olarak kullanılıyor olması,

2. Kökeninin tarih öncesi/kadim dönemlere dayanıp, İslamiyet öncesi ve sonrası dönemlerde sembolik anlam sürekliliği göstermesi,

3. Yapay zekâ metin-görsel (text-to-image) modellerinin geometrik ve biçimsel olarak algılayabileceği belirgin bir taksonomik/grafik forma sahip olması.

Bu kriterler doğrultusunda literatür taraması yapılmış ve Türk dünyasının ortak belleğini en güçlü şekilde temsil eden [Koçboynuzu, Hayat Ağacı, Çift Başlı Kartal, Eli Belinde, KünAy -Güneş ve Yıldız, Geometrik Geçmeler ve Yıldız Formları (Girih / Selçuklu Yıldızı), Kurt (Bozkurt) Sembolü] motifleri çalışma kapsamına alınmıştır. Seçilen motifler, biçimsel (formal) ve anlamsal (semantic) özellikleri bakımından kodlanarak yapay zekâ komut (prompt) matrislerine dönüştürülmüştür.

B. Görsel Üretim ve Yapay Zekâ Parametreleri (Uygulama)

Görselleştirme sürecinde, güncel difüzyon modelleri arasında sembolik ve sanatsal üretim kapasitesi yüksek olan OpenAI tarafından geliştirilen DALL-E 3 (ChatGPT tabanlı) yapay zekâ motoru kullanılmıştır. Kültürel önyargı (cultural bias) faktörünü en aza indirmek ve motiflerin özgün geometrisini korumak adına prompt mühendisliği (prompt engineering) tekniklerinden yararlanılmıştır. Promptlar kurgulanırken sadece motif isimleri değil; motiflerin matematiksel simetrisi, stilizasyon ilkeleri, geleneksel renk paletleri ve modern tasarım disiplinlerine (bu çalışmada amblem tasarımı) uyarlanma biçimleri teknik parametrelerle standartlaştırılarak sisteme girilmiştir.

C. Sınırlılıklar (Limitations)

Her bilimsel çalışmada olduğu gibi bu araştırmanın da belirli sınırlılıkları bulunmaktadır:

• Coğrafi Sınırlılık: Çalışma, Türk dünyasının zengin çeşitliliğine sahip özerk cumhuriyetlerini ve topluluklarını (Gagavuzya, Tataristan, Tuva vb.) kapsamayıp, yalnızca altı bağımsız Türk devleti ile sınırlandırılmıştır.

• Tipolojik Sınırlılık: Türk görsel kültüründeki yüzlerce varyasyon arasından yalnızca ortak niteliği en yüksek olan [7] adet motif analize dâhil edilmiştir.

• Teknolojik Sınırlılık (Veri Seti Ön Yargısı): Kullanılan yapay zekâ modellerinin ağırlıklı olarak Batı merkezli veri setleri (Western-centric datasets) ile eğitilmiş olması, Türk dünyasına ait bazı kadim sembolleri yorumlarken deformasyona veya kültürel sapmaya yol açabilmektedir. Bu durum, üretilen görsellerin yapay zekânın 'kendi yorum kapasitesiyle' sınırlı olduğunu ortaya koymaktadır."

VI. BULGULAR

A. Türk Dünyası Ortak Sembollerinin Dijital Dönüşümü ve Amblem Tasarımına Yansımaları

Araştırmanın bulgular bölümü, belirlenen yöntem dâhilinde OpenAI DALL-E 3 yapay zekâ modeline aktarılan taksonomik prompt matrislerinin görsel çıktılarından ve bu çıktıların göstergebilimsel/biçimsel analizinden oluşmaktadır. Bu doğrultuda; Türk dünyasının ortak görsel belleğini en güçlü şekilde temsil eden yedi temel motif (Koçboynuzu, Hayat Ağacı, Çift Başlı Kartal, Eli Belinde, KünAy, Geometrik Geçmeler ve Girih/Selçuklu Yıldızı, Kurt Sembolü), kurumsal kimlik ve amblem tasarımı disiplininin ilkelerine uygun olarak yeniden kurgulanmıştır. Görsel üretim sürecinde, sembollerin monokromatik ve yalın form bütünlüğünü test etmek amacıyla siyah zemin (fon) tercih edilmiş; tasarımlara zamansızlık, prestij ve kurumsal temsiliyet kazandırmak adına metalik (altın, gümüş ve bronz) renk paletleri uygulanmıştır. Bulgular, kadim sembollerin hem geleneksel yüzey şemalarını koruyan 2 boyutlu (2D) minimalist varyasyonlarını hem de modern dijital ekranlar ve mekânsal uygulamalar için derinlik/malzeme algısı yüksek 3 boyutlu (3D) amblem formlarını içermektedir. Elde edilen çıktılar, yapay zekânın Batı merkezli veri seti ön yargısına rağmen, doğru prompt mühendisliği teknikleriyle Türk görsel kültürünün geometrik simetrisini, stilizasyon ilkelerini ve sembolik dilini deformasyona uğratmadan amblem formunda somutlaştırabildiğini göstermektedir.

1. Koçboynuzu Motifi

Türk dünyasının en yaygın ortak sembollerinden biridir. Güç, erkeklik, bereket, kahramanlık ve koruyuculuğu temsil eder.



Fig. 1. Koç Boynuzu Sembolü Yapay Zeka Uyarlaması

Üretilen tasarımlarda motif, geleneksel biçimsel karakterini korurken çağdaş grafik tasarım anlayışına uygun olarak sadeleştirilmiş, simetrik ve kurumsal kimlik çalışmalarında kullanılabilecek bir amblem yapısına dönüştürülmüştür.

Elde edilen tasarımlar incelendiğinde, koç boynuzu motifinin kıvrımlı yapısının vektörel çizgiler aracılığıyla yeniden yorumlandığı görülmektedir. Tasarımlarda özellikle eksenel simetri prensibi kullanılarak denge, bütünlük ve görsel istikrar sağlanmıştır. Geleneksel motifin spiral karakteri korunurken, geometrik çizgilerle desteklenen modern bir form dili geliştirilmiştir. Bu yaklaşım, kültürel bir sembolün günümüz görsel iletişim ihtiyaçlarına uyarlanabileceğini göstermektedir.

İki boyutlu (2D) tasarımlar, sade yüzey kullanımı, net konturlar ve yüksek okunabilirlik özellikleriyle öne çıkmaktadır. Bu tasarımlar, logo ve kurumsal kimlik uygulamalarında ölçeklenebilirlik ve işlevsellik açısından avantaj sağlamaktadır. Minimal çizgi yapısı sayesinde motifin sembolik anlamı korunurken çağdaş bir görsel kimlik oluşturulmuştur.



Fig. 2. Koç Boynuzu Sembolü 3D Yapay Zeka Uyarlaması

İkinci grupta yer alan üç boyutlu (3D) tasarımlar ise metalik yüzey etkileri, ışık-gölge geçişleri ve derinlik algısı kullanılarak geliştirilmiştir. Altın, bronz ve gümüş tonlarında oluşturulan yüzeyler, motife daha güçlü bir görsel etki kazandırmış; geleneksel sembolün prestij, güç ve değer kavramlarıyla ilişkilendirilmesini desteklemiştir. Üç boyutlu yorumlarda koç boynuzu formu yalnızca dekoratif bir unsur olmaktan çıkarılarak dijital ortamda dikkat çekici ve marka değeri yüksek bir amblem niteliğine dönüştürülmüştür.

Görsel analiz sonucunda tasarımların tamamında koç boynuzu motifinin kültürel çağrışımlarının korunduğu tespit edilmiştir. Türk kültüründe güç, cesaret, bereket, koruyuculuk ve kahramanlık gibi anlamlarla ilişkilendirilen motif, yapay zekâ tarafından üretilen bu tasarımlarda da sembolik işlevini sürdürmektedir. Bununla birlikte yapay zekâ araçlarının, geleneksel motifleri farklı biçimsel varyasyonlarla yeniden üreterek tasarımcıya çok sayıda alternatif sunabildiği gözlemlenmiştir. Bu durum, kültürel miras unsurlarının dijital

tasarım süreçlerine aktarılmasında yapay zekânın önemli bir yardımcı araç olarak kullanılabileceğini göstermektedir.

2. Hayat Ağacı Motifi

Yaşamın sürekliliği, soy bağı, evren düzeni ve ruhsal yükselişi temsil eder. Halı ve kilimler, minyatürler, ahşap oymalar, mimari süslemeler, nakış ve suzani işlerinde örnekleri görülür.

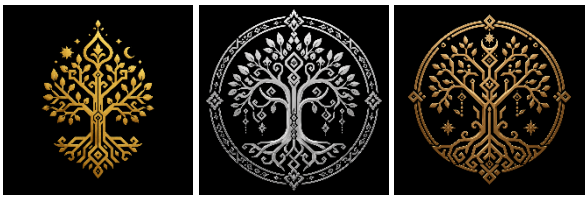


Fig. 3. Hayat Ağacı Sembolü Yapay Zeka Uyarlaması

Üretilen tasarımlarda hayat ağacı motifi, geleneksel biçimsel karakterini ve köklü sembolik yapısını korurken çağdaş grafik tasarım anlayışına uygun olarak sadeleştirilmiş, simetrik ve kurumsal kimlik çalışmalarında kullanılabilecek dinamik bir amblem yapısına dönüştürülmüştür. Geleneksel bezeme sanatları bünyesinde sıkça rastlanan dikey yerleşimli ve dallanan kompozisyon şeması, yapay zekâ üretim süreçlerinde dijital yüzey geometrisiyle optimize edilerek modern amblem tasarım ilkelerine uyarlanmıştır.

Elde edilen tasarımlar incelendiğinde, hayat ağacı motifinin kök, gövde, dal ve yaprak unsurlarından oluşan organik yapısının vektörel çizgiler ve net konturlar aracılığıyla yeniden yorumlandığı görülmektedir. Tasarımlarda özellikle eksenel simetri prensibi kullanılarak denge, bütünlük, optik merkezleme ve görsel istikrar sağlanmıştır. Geleneksel motifin büyüme, yükselme ve dallanma karakteri korunurken; dairesel, rozet formlu veya dikey akslı geometrik çizgilerle desteklenen modern bir form dili geliştirilmiştir. Bu yaklaşım, evrensel ve kültürel bir sembolün günümüz görsel iletişim ihtiyaçlarına, logo, piktogram ve ikonografi standartlarına başarıyla uyarlanabileceğini göstermektedir.

İki boyutlu (2D) tasarımlar; sade yüzey kullanımı, net çizgisel hatlar, negatif-pozitif alan dengesi ve yüksek okunabilirlik (legibility) özellikleriyle öne çıkmaktadır. Bu tasarımlar, logo ve kurumsal kimlik uygulamalarında, farklı yüzey ve boyutlarda ölçeklenebilirlik ile işlevsellik açısından üst düzey avantaj sağlamaktadır. Minimal ve geometrik çizgi yapısı sayesinde motifin sembolik derinliği rafine edilirken, çağdaş bir görsel kimlik dili oluşturulmuştur.



Fig. 4. Hayat Ağacı Sembolü 3D Yapay Zeka Uyarlaması

İkinci grupta yer alan üç boyutlu (3D) tasarımlar ise metalik yüzey etkileri, hassas ışık-gölge geçişleri, gradyanlar ve yüksek derinlik algısı kullanılarak geliştirilmiştir. Altın, gümüş ve bronz tonlarında oluşturulan yüzey dokuları, motife daha güçlü, prestijli ve estetik bir görsel etki kazandırmış; geleneksel sembolün kurumsal temsiliyet, süreklilik, kalite ve değer kavramlarıyla ilişkilendirilmesini desteklemiştir. Üç boyutlu yorumlarda hayat ağacı formu yalnızca dekoratif bir bezeme unsuru olmaktan çıkarılarak, dijital mecra ve çağdaş markalama süreçlerinde dikkat çekici, plastik değeri yüksek ve marka algısı güçlü bir amblem niteliğine dönüştürülmüştür.

Görsel analiz sonucunda tasarımların tamamında hayat ağacı motifinin kültürel çağrışımlarının ve arkaik anlam katmanlarının başarıyla korunduğu tespit edilmiştir. Türk kültüründe ve mitolojisinde evrenin düzeni, doğum, yaşamın sürekliliği, ölümsüzlük, soyun devamlılığı ve kozmik merkez (axis mundi) gibi derin anlamlarla ilişkilendirilen motif, yapay zekâ tarafından üretilen bu tasarımlarda da sembolik işlevini ve görsel gücünü eksiksiz bir biçimde sürdürmektedir. Bununla birlikte yapay zekâ araçlarının, geleneksel motifleri kök-dal ilişkisi, yaprak yoğunluğu ve çerçeve geometrisi gibi farklı biçimsel varyasyonlarla yeniden üreterek tasarımcıya çok kısa sürede çok sayıda nitelikli alternatif sunabildiği gözlemlenmiştir. Bu durum, kültürel miras unsurlarının dijital tasarım süreçlerine, amblem ve kurumsal kimlik alanlarına aktarılmasında yapay zekânın yaratıcı süreci besleyen ve tasarım ekosistemini zenginleştiren önemli bir yardımcı araç olarak kullanılabileceğini ortaya koymaktadır.

3. Çift Başlı Kartal Motifi

Çift başlı kartal; Türk dünyasında güç, hâkimiyet, koruyuculuk ve göksel otoriteyi temsil eden önemli ortak sembollerden biridir. Özellikle Türkiye merkezli Anadolu Selçuklu Devleti sanatında belirginleşmiş, ancak Orta Asya Türk kültürlerinde de farklı biçimlerde varlığını sürdürmüştür.

Güç ve kudret, hükümdarlık, doğu ve Batı'ya hâkimiyet koruyuculuk, göksel otorite, ruhsal ve dünyevi denge anlamlarında kullanılır.



Fig. 5. Çift Başlı Kartal Sembolü Yapay Zeka Uyarlaması

Üretilen tasarımlarda motif, geleneksel biçimsel karakterini korurken çağdaş grafik tasarım anlayışına uygun olarak sadeleştirilmiş, simetrik ve kurumsal kimlik çalışmalarında kullanılabilecek bir amblem yapısına dönüştürülmüştür. Geleneksel Türk sanatlarında ve mimarisinde ikonik bir yer tutan çift başlı yerleşim planı, yapay zekâ üretim süreçlerinde dijital yüzey geometrisiyle optimize edilerek modern amblem tasarım ilkelerine uyarlanmıştır.

Tasarımlar incelendiğinde, çift başlı kartal motifinin açılmış kanat, gövde, kuyruk ve zıt yönlere bakan baş unsurlarından oluşan görkemli yapısının vektörel çizgiler aracılığıyla yeniden yorumlandığı görülmektedir. Tasarımlarda özellikle eksenel simetri prensibi kullanılarak denge, bütünlük ve görsel istikrar sağlanmıştır. Geleneksel motifin ikili/simetrik karakteri korunurken, gövde ve kanat alanlarında geometrik çizgiler ile iç içe geçen örgü motifleriyle desteklenen modern bir form dili geliştirilmiştir. Bu yaklaşım, kültürel bir sembolün günümüz görsel iletişim ihtiyaçlarına uyarlanabileceğini göstermektedir.

İki boyutlu (2D) tasarımlar, sade yüzey kullanımı, net konturlar ve yüksek okunabilirlik özellikleriyle öne çıkmaktadır. Bu tasarımlar, logo ve kurumsal kimlik uygulamalarında ölçeklenebilirlik ve işlevsellik açısından avantaj sağlamaktadır. Minimal çizgi yapısı ve stilize edilmiş tüyler sayesinde motifin sembolik anlamı korunurken çağdaş bir görsel kimlik oluşturulmuştur.



Fig. 6. Çift Başlı Kartal Sembolü 3D Yapay Zeka Uyarlaması

İkinci grupta yer alan üç boyutlu (3D) tasarımlar ise metalik yüzey etkileri, ışık-gölge geçişleri ve derinlik algısı kullanılarak geliştirilmiştir. Altın, bronz ve gümüş tonlarında oluşturulan yüzeyler, motife daha güçlü bir görsel etki kazandırmış; geleneksel sembolün prestij, güç ve değer kavramlarıyla ilişkilendirilmesini desteklemiştir. Üç boyutlu yorumlarda çift başlı kartal formu yalnızca dekoratif bir unsur olmaktan çıkarılarak dijital ortamda dikkat çekici ve marka değeri yüksek bir amblem niteliğine dönüştürülmüştür.

Görsel analiz sonucunda tasarımların tamamında çift başlı kartal motifinin kültürel çağrışımlarının korunduğu tespit edilmiştir. Türk kültüründe, Selçuklu devlet geleneğinde ve mitolojisinde güç, egemenlik, çift yönlü hâkimiyet (doğu-batı), koruyuculuk ve asalet gibi anlamlarla ilişkilendirilen motif, yapay zekâ tarafından üretilen bu tasarımlarda da sembolik işlevini sürdürmektedir. Bununla birlikte yapay zekâ araçlarının, geleneksel motifleri kanat açısı, tüy detayları ve tepe sembolleri (ay, yıldız, geometrik bezemeler) gibi farklı biçimsel varyasyonlarla yeniden üreterek tasarımcıya çok sayıda alternatif sunabildiği gözlemlenmiştir. Bu durum, kültürel miras unsurlarının dijital tasarım süreçlerine aktarılmasında yapay zekânın önemli bir yardımcı araç olarak kullanılabileceğini göstermektedir.

4. Eli Belinde Motifi

Bereket, doğurganlık, ana tanrıça ve kadınlığı temsil eder. Anadolu, Türkmenistan, Kırgız ve Kazak dokumalarında sıklıkla görülen bir motiftir.

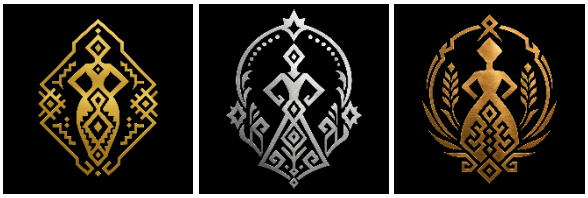


Fig. 7. Eli Belinde Sembolü Yapay Zeka Uyarlaması

Üretilen tasarımlarda motif, geleneksel biçimsel karakterini korurken çağdaş grafik tasarım anlayışına uygun olarak sadeleştirilmiş, simetrik ve kurumsal kimlik çalışmalarında kullanılabilecek bir amblem yapısına dönüştürülmüştür. Geleneksel dokuma ve el sanatlarında dişiliğin, doğurganlığın ve soyun devamlılığının simgesi olan antropomorfik yerleşim şeması, yapay zekâ üretim süreçlerinde dijital yüzey geometrisiyle optimize edilerek modern amblem tasarım ilkelerine uyarlanmıştır.

Tasarımlar incelendiğinde, eli belinde motifinin kollarını yanlara açarak gövdesine dayayan şematize kadın figürü yapısının vektörel çizgiler aracılığıyla yeniden yorumlandığı görülmektedir. Tasarımlarda özellikle eksenel simetri prensibi kullanılarak denge, bütünlük ve görsel istikrar sağlanmıştır. Geleneksel motifin köşeli ve geometrik karakteri korunurken, gövde içlerindeki pötrak veya baklava bezemeleriyle desteklenen modern bir form dili geliştirilmiştir. Bu yaklaşım, kültürel bir sembolün günümüz görsel iletişim ihtiyaçlarına uyarlanabileceğini göstermektedir.

İki boyutlu (2D) tasarımlar, sade yüzey kullanımı, net konturlar ve yüksek okunabilirlik özellikleriyle öne çıkmaktadır. Bu tasarımlar, logo ve kurumsal kimlik uygulamalarında ölçeklenebilirlik ve işlevsellik açısından avantaj sağlamaktadır. Minimal çizgi yapısı ve geometrik hatlar sayesinde motifin sembolik anlamı korunurken çağdaş bir görsel kimlik oluşturulmuştur.



Fig. 8. Eli Belinde Sembolü 3D Yapay Zeka Uyarlaması

İkinci grupta yer alan üç boyutlu (3D) tasarımlar ise metalik yüzey etkileri, ışık-gölge geçişleri ve derinlik algısı kullanılarak geliştirilmiştir. Altın, gümüş ve bronz tonlarında oluşturulan yüzeyler, motife daha güçlü bir görsel etki kazandırmış; geleneksel sembolün prestij, güç ve değer kavramlarıyla ilişkilendirilmesini desteklemiştir. Üç boyutlu yorumlarda eli belinde formu yalnızca dekoratif bir unsur olmaktan çıkarılarak dijital ortamda dikkat çekici ve marka değeri yüksek bir amblem niteliğine dönüştürülmüştür.

Görsel analiz sonucunda tasarımların tamamında eli belinde motifinin kültürel çağrışımlarının korunduğu tespit edilmiştir. Türk kültüründe, dokuma literatüründe ve Anadolu sembolizminde bereket, ana tanrıça kültü, kısmet, neşe ve üreme gibi anlamlarla ilişkilendirilen motif, yapay zekâ tarafından üretilen bu tasarımlarda da sembolik işlevini sürdürmektedir. Bununla birlikte yapay zekâ araçlarının, geleneksel motifleri dış çerçeve geometrisi (baklava, dairesel madalyon), baş üstü sembolleri (hilal, yıldız) ve çevreleyici bitkisel başak formları gibi farklı biçimsel varyasyonlarla yeniden üreterek tasarımcıya çok sayıda alternatif sunabildiği gözlemlenmiştir. Bu durum, kültürel miras unsurlarının dijital tasarım süreçlerine aktarılmasında yapay zekânın önemli bir yardımcı araç olarak kullanılabileceğini göstermektedir.

5. Kün-ay, Güneş ve Yıldız Motifleri

Kozmik düzen, kutsallık, yön bulma ve devlet simgesi olarak kullanılır. Bayraklar, Kubbe süslemeleri, Takılar, Çadır dekorları vb. yaygın kullanım alanlarıdır.



Fig. 9. Kün-ay, Güneş ve Yıldız Sembolü Yapay Zeka Uyarlaması

Üretilen tasarımlarda motif, geleneksel biçimsel karakterini korurken çağdaş grafik tasarım anlayışına uygun olarak sadeleştirilmiş, simetrik ve kurumsal kimlik çalışmalarında kullanılabilecek bir amblem yapısına dönüştürülmüştür. Geleneksel Türk kozmolojisinde ve ikonografisinde merkezi bir yer tutan göksel sembollerin yerleşim planı, yapay zekâ üretim süreçlerinde dijital yüzey geometrisiyle optimize edilerek modern amblem tasarım ilkelerine uyarlanmıştır.

Tasarımlar incelendiğinde, kün-ay, güneş ve yıldız motiflerinin merkezcil ışınlar, dairesel hatlar ve çok köşeli yıldız formlarından oluşan geometrik yapısının vektörel çizgiler aracılığıyla yeniden yorumlandığı görülmektedir. Tasarımlarda özellikle radyal ve eksenel simetri prensipleri bir arada kullanılarak denge, bütünlük ve görsel istikrar sağlanmıştır. Geleneksel motifin göksel ve döngüsel karakteri korunurken, keskin ışınsal hatlar ve eş merkezli dairesel çizgilerle desteklenen modern bir form dili geliştirilmiştir. Bu yaklaşım, kültürel bir sembolün günümüz görsel iletişim ihtiyaçlarına uyarlanabileceğini göstermektedir.

İki boyutlu (2D) tasarımlar, sade yüzey kullanımı, net konturlar ve yüksek okunabilirlik özellikleriyle öne çıkmaktadır. Bu tasarımlar, logo ve kurumsal kimlik uygulamalarında ölçeklenebilirlik ve işlevsellik açısından avantaj sağlamaktadır. Minimal çizgi yapısı ve geometrik düzen sayesinde motifin sembolik anlamı korunurken çağdaş bir görsel kimlik oluşturulmuştur.

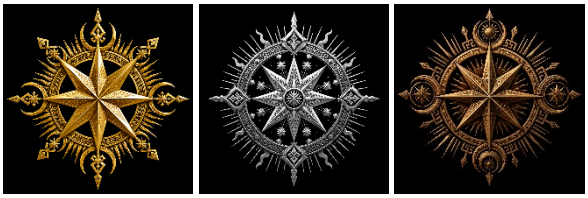


Fig. 10. Kün-ay, Güneş ve Yıldız Sembolü 3D Yapay Zeka Uyarlaması

İkinci grupta yer alan üç boyutlu (3D) tasarımlar ise metalik yüzey etkileri, ışık-gölge geçişleri ve derinlik algısı kullanılarak geliştirilmiştir. Altın, bronz ve gümüş tonlarında oluşturulan yüzeyler, motife daha güçlü bir görsel etki kazandırmış; geleneksel sembolün prestij, güç ve değer kavramlarıyla ilişkilendirilmesini desteklemiştir. Üç boyutlu yorumlarda kün-ay, güneş ve yıldız formları yalnızca dekoratif bir unsur olmaktan çıkarılarak dijital ortamda dikkat çekici ve marka değeri yüksek bir amblem niteliğine dönüştürülmüştür.

Görsel analiz sonucunda tasarımların tamamında kün-ay, güneş ve yıldız motiflerinin kültürel çağrışımlarının korunduğu tespit edilmiştir. Türk kültüründe, erken dönem inançlarında ve devlet sembolizminde kozmik düzen, yaşam kaynağı, aydınlık, talih, egemenlik ve zamanın döngüsü gibi anlamlarla ilişkilendirilen bu motifler, yapay zekâ tarafından üretilen bu tasarımlarda da sembolik işlevini sürdürmektedir. Bununla birlikte yapay zekâ araçlarının, geleneksel motifleri köşe sayıları, ışın yoğunlukları, dairesel katmanlar ve hilal-yıldız kombinasyonları gibi farklı biçimsel varyasyonlarla yeniden üreterek tasarımcıya çok sayıda alternatif sunabildiği gözlemlenmiştir. Bu durum, kültürel miras unsurlarının dijital tasarım süreçlerine aktarılmasında yapay zekânın önemli bir yardımcı araç olarak kullanılabileceğini göstermektedir.

6. Geometrik Geçmeler ve Yıldız Formları (Girih / Selçuklu Yıldızı) Motifleri

Geometrik geçmeler ve yıldız kompozisyonları; Türk dünyasında matematiksel düzeni, sonsuzluğu, evrenin harmonisini ve birlik anlayışını temsil eden en yaygın ortak görsel sistemlerden biridir. Özellikle Türkiye, Azerbaycan, Özbekistan, Türkmenistan, Kazakistan ve Kırgızistan kültürlerinde mimari, çini, halı ve süsleme sanatlarında yoğun biçimde görülür.

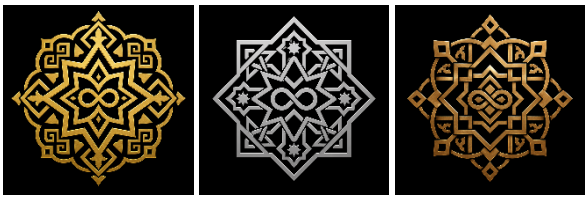


Fig. 11. Geometrik Geçmeler ve Yıldız Formları Sembolü Yapay Zeka Uyarlaması

Üretilen tasarımlarda motif, geleneksel biçimsel karakterini korurken çağdaş grafik tasarım anlayışına uygun olarak sadeleştirilmiş, simetrik ve kurumsal kimlik çalışmalarında kullanılabilecek bir amblem yapısına dönüştürülmüştür. Geleneksel İslam ve Türk mimari süsleme sanatlarında merkezi bir yer tutan çokgen esaslı düğüm ve geçme düzenleri, yapay zekâ üretim süreçlerinde dijital yüzey geometrisiyle optimize edilerek modern amblem tasarım ilkelerine uyarlanmıştır.

Görselde yer alan tasarımlar incelendiğinde, girih ve Selçuklu yıldızı motiflerinin iç içe geçen şeritler, kesişen doğrusal hatlar ve merkezkaç çok kollu yıldız yapılarından oluşan matematiksel kompozisyonunun vektörel çizgiler aracılığıyla yeniden yorumlandığı görülmektedir. Tasarımlarda özellikle eksenel ve radyal simetri prensipleri bir arada kullanılarak denge, bütünlük ve görsel istikrar sağlanmıştır. Geleneksel motifin sonsuzluk hissi uyandıran geçme karakteri korunurken, geometrik çizgilerle desteklenen modern bir form dili geliştirilmiştir. Bu yaklaşım, kültürel bir sembolün günümüz görsel iletişim ihtiyaçlarına uyarlanabileceğini göstermektedir.

İki boyutlu (2D) tasarımlar, sade yüzey kullanımı, net konturlar ve yüksek okunabilirlik özellikleriyle öne çıkmaktadır. Bu tasarımlar, logo ve kurumsal kimlik uygulamalarında ölçeklenebilirlik ve işlevsellik açısından avantaj sağlamaktadır. Minimal çizgi yapısı ve keskin geometrik hatlar sayesinde motifin sembolik anlamı korunurken çağdaş bir görsel kimlik oluşturulmuştur.



Fig. 12. Geometrik Geçmeler ve Yıldız Formları Sembolü 3D Yapay Zeka Uyarlaması

İkinci grupta yer alan üç boyutlu (3D) tasarımlar ise metalik yüzey etkileri, ışık-gölge geçişleri ve şeritlerin alt-üst kesişim noktalarındaki derinlik algısı kullanılarak geliştirilmiştir. Altın, bronz ve gümüş tonlarında oluşturulan yüzeyler, motife daha güçlü bir görsel etki kazandırmış; geleneksel sembolün prestij, güç ve değer kavramlarıyla

ilişkilendirilmesini desteklemiştir. Üç boyutlu yorumlarda Selçuklu yıldızı ve girih formu yalnızca dekoratif bir unsur olmaktan çıkarılarak dijital ortamda dikkat çekici ve marka değeri yüksek bir amblem niteliğine dönüştürülmüştür.

Görsel analiz sonucunda tasarımların tamamında geometrik geçme ve yıldız motiflerinin kültürel çağrışımlarının korunduğu tespit edilmiştir. Türk kültüründe, Selçuklu mimari mirasında ve sanat felsefesinde evrenin kozmik düzeni, adalet, eşitlik, bilgelik ve sonsuzluk gibi derin anlamlarla ilişkilendirilen motif, yapay zekâ tarafından üretilen bu tasarımlarda da sembolik işlevini sürdürdüğünü göstermektedir. Bununla birlikte yapay zekâ araçlarının, geleneksel motifleri köşe sayıları, geçme örüntülerinin yoğunluğu, merkezdeki düğüm varyasyonları ve dış çerçeve kurguları gibi farklı biçimsel alternatiflerle yeniden üreterek tasarımcıya çok sayıda seçenek sunabildiği gözlemlenmiştir. Bu durum, kültürel miras unsurlarının dijital tasarım süreçlerine aktarılmasında yapay zekânın önemli bir yardımcı araç olarak kullanılabileceğini göstermektedir.

7. Kurt (Bozkurt) Sembolü

Türk mitolojisinin en önemli figürlerinden biridir. Rehberlik, bağımsızlık ve soyun kökenini temsil eder. Ergenekon Destanı, Asena miti, Göktürk sembolizmi vb. temsil eder. Arma ve logolar, heykeller, çağdaş grafik tasarımlar yoğun olarak kullanıldığı alanlardır.



Fig. 13. Kurt (Bozkurt) Sembolü Yapay Zeka Uyarlaması

Üretilen tasarımlarda motif, geleneksel biçimsel karakterini korurken çağdaş grafik tasarım anlayışına uygun olarak sadeleştirilmiş, simetrik ve kurumsal kimlik çalışmalarında kullanılabilecek bir amblem yapısına dönüştürülmüştür. Geleneksel Türk mitolojisinde ve destan kültüründe yol göstericiliğin, bağımsızlığın ve askeri gücün sembolü olan kurt figürünün yerleşim planı, yapay zekâ üretim süreçlerinde dijital yüzey geometrisiyle optimize edilerek modern amblem tasarım ilkelerine uyarlanmıştır.

Tasarımlar incelendiğinde, kurt motifinin dik kulaklar, karakteristik yüz hatları ve cepheden keskin bakış açısıyla kurgulanan organik yapısının vektörel çizgiler aracılığıyla yeniden yorumlandığı görülmektedir. Tasarımlarda özellikle eksenel simetri prensibi kullanılarak denge, bütünlük ve görsel istikrar sağlanmıştır. Geleneksel motifin dinamik ve mağrur karakteri korunurken; dairesel hatlar, yarım ay formları ve Göktürk dönemi runik harfleriyle/tamgalarıyla desteklenen modern bir form dili geliştirilmiştir. Bu yaklaşım, kültürel bir sembolün günümüz görsel iletişim ihtiyaçlarına uyarlanabileceğini göstermektedir.

İki boyutlu (2D) tasarımlar, sade yüzey kullanımı, net konturlar ve yüksek okunabilirlik özellikleriyle öne çıkmaktadır. Bu tasarımlar, logo ve kurumsal kimlik uygulamalarında ölçeklenebilirlik ve işlevsellik açısından avantaj sağlamaktadır. Minimal çizgi yapısı ve stilize edilmiş kürk detayları sayesinde motifin sembolik anlamı korunurken çağdaş bir görsel kimlik oluşturulmuştur.



Fig. 14. Kurt (Bozkurt) Sembolü 3D Yapay Zeka Uyarlaması

İkinci grupta yer alan üç boyutlu (3D) tasarımlar ise metalik yüzey etkileri, ışık-gölge geçişleri ve kabartma derinliği algısı kullanılarak geliştirilmiştir. Altın, gümüş ve bronz tonlarında oluşturulan yüzeyler, motife daha güçlü bir görsel etki kazandırmış; geleneksel sembolün prestij, güç ve değer kavramlarıyla ilişkilendirilmesini desteklemiştir. Üç boyutlu yorumlarda kurt formu yalnızca dekoratif bir unsur olmaktan çıkarılarak dijital ortamda dikkat çekici ve marka değeri yüksek bir amblem niteliğine dönüştürülmüştür.

Görsel analiz sonucunda tasarımların tamamında kurt motifinin kültürel çağrışımlarının korunduğu tespit edilmiştir. Türk kültüründe, epik destanlarda ve devlet geleneğinde hürriyet, cesaret, çeviklik, koruyuculuk ve rehberlik gibi anlamlarla ilişkilendirilen motif, yapay zekâ tarafından üretilen bu tasarımlarda da sembolik işlevini sürdürmektedir. Bununla birlikte yapay zekâ araçlarının, geleneksel motifleri çevreleyici ay-yıldız kurguları, arka plan geometrisi, runik yazı entegrasyonu ve tam profilden cepheye uzanan anatomik açılar gibi farklı biçimsel varyasyonlarla yeniden üreterek tasarımcıya çok sayıda alternatif sunabildiği gözlemlenmiştir. Bu durum, kültürel miras unsurlarının dijital tasarım süreçlerine aktarılmasında yapay zekânın önemli bir yardımcı araç olarak kullanılabileceğini göstermektedir.

VII. TARTIŞMA SONUÇ VE ÖNERİLER

Bu araştırma, Türk dünyasının ortak görsel kültür mirasında yer alan geleneksel motiflerin üretken yapay zekâ teknolojileri aracılığıyla yeniden yorumlanabileceğini ve çağdaş tasarım uygulamalarına aktarılabileceğini ortaya koymuştur. Araştırma kapsamında incelenen Koçboynuzu, Hayat Ağacı, Çift Başlı Kartal, Eli Belinde, Kün-Ay, Geometrik Geçmeler (Girih) ve Bozkurt motiflerinin, yapay zekâ destekli görselleştirme süreçlerinde biçimsel bütünlüklerini büyük ölçüde koruyarak çağdaş amblem ve kurumsal kimlik tasarımı diline uyarlanabildiği görülmüştür. Elde edilen sonuçlar, bu sembollerin yalnızca tarihsel ve etnografik birer kültürel unsur olarak değil, günümüz görsel iletişim ortamlarında kullanılabilecek dinamik tasarım bileşenleri olarak da değerlendirilebileceğini göstermektedir.

Araştırma bulguları, yapay zekânın yalnızca teknik bir görsel üretim aracı olmadığını; aynı zamanda kültürel mirasın görünür kılınması, yeniden dolaşıma sokulması ve yeni kuşaklarla buluşturulması açısından önemli bir potansiyel taşıdığını ortaya koymaktadır. Dijitalleşme sürecinin yalnızca koruma ve arşivleme faaliyetlerinden ibaret olmadığı, kültürel bilginin yeniden üretilmesine ve yaratıcı endüstrilere aktarılmasına da olanak tanıdığı anlaşılmaktadır. Bu bağlamda yapay zekâ destekli tasarım süreçleri, geleneksel kültürel değerlerin güncel iletişim ortamlarında yeniden temsil edilmesine katkı sağlayarak kültürel devamlılığı destekleyen yeni üretim modelleri sunmaktadır.

Literatürde yapay zekânın sanatsal üretim süreçlerine dâhil olması; özgünlük, yaratıcı niyet, estetik değer ve

zanaatkârlık bağlamında önemli tartışmaları beraberinde getirmektedir. Yapay zekâ destekli üretimlerin geleneksel sanat pratiğini ve özgün yaratım süreçlerini zayıflattığı yönündeki eleştirilere rağmen, bu araştırmada uygulanan sistematik prompt mühendisliği yaklaşımı, yapay zekânın doğru yönlendirmelerle Türk estetik düşüncesinin temelini oluşturan simetri, stilizasyon ve geometrik düzen ilkelerini koruyabildiğini göstermiştir. Böylece tasarım sürecinde fiziksel üretim becerisinin yerini giderek daha fazla kavramsal tasarım, küratöryel yaklaşım ve yaratıcı yönlendirme süreçlerinin aldığı görülmektedir.

Araştırmada elde edilen amblem tasarımlarının minimalist, monokromatik ve metalik estetik anlayışla yeniden yorumlanması, geleneksel sembollerin statik kültürel nesneler olmaktan çıkarılarak çağdaş marka kimliği, kurumsal iletişim ve dijital tasarım alanlarında yeniden üretilebilir görsel unsurlara dönüştürülebileceğini ortaya koymuştur. Bu durum, insan ve yapay zekâ arasındaki iş birliğine dayalı yaratıcı üretim modellerinin kültürel tasarım süreçlerinde uygulanabilirliğini desteklemektedir.

Çalışmanın önemli bulgularından biri de yapay zekâ sistemlerinin kültürel veri kümelerine bağımlı olarak çalışması ve bu veri kümelerinde bulunan temsil eksikliklerinin üretilen görsellere doğrudan yansımasıdır. Özellikle Batı merkezli veri setleriyle eğitilmiş üretken yapay zekâ modellerinin bazı Türk kültürü sembollerini yanlış yorumlayabildiği, Batı heraldik geleneklerine veya popüler kültür imgelerine yaklaştırabildiği gözlemlenmiştir. Özellikle Bozkurt, Çift Başlı Kartal ve Kün-Ay gibi sembollerin ilk üretimlerde kültürel bağlamlarından uzaklaşabildiği görülmüştür. Ancak ayrıntılı semantik etiketleme, kültürel referansların güçlendirilmesi ve taksonomik prompt mühendisliği teknikleri sayesinde bu sapmaların önemli ölçüde azaltılabildiği belirlenmiştir. Bu durum, kültürel miras odaklı yapay zekâ çalışmalarında yerel ve kültüre özgü veri setlerinin geliştirilmesinin önemini açık biçimde ortaya koymaktadır.

Araştırma sonucunda aşağıdaki temel değerlendirmelere ulaşılmıştır:

1. Türk dünyasına ait geleneksel motifler, taşıdıkları güçlü sembolik anlamlar ve geometrik yapılar sayesinde üretken yapay zekâ sistemleri tarafından yüksek doğrulukla analiz edilebilmekte ve çağdaş grafik tasarım ilkeleri doğrultusunda yeniden yorumlanabilmektedir.

2. Yapay zekâ ile kültürel mirasın yeniden üretilmesi süreci yalnızca teknik bir dönüşüm değil, kültürel sürdürülebilirliği ve kolektif hafızanın dijital ortamda yeniden inşasını destekleyen yaratıcı bir paradigma olarak değerlendirilebilir.

3. Küresel yapay zekâ modellerinde görülen Batı merkezli veri seti kısıtlılıkları, Türk dünyasına ait bazı sembollerin temsilinde sorunlara yol açabilmekte; bu durum ancak yerel veri kümeleri ve kültürel bağlamı güçlendirilmiş tasarım süreçleriyle aşılabilmektedir.

4. Geleneksel motiflerin çağdaş tasarım ortamlarında yeniden kullanılması, kültürel kimliğin görünürlüğünü artırmakta ve kültürel mirasın dijital çağda sürdürülebilirliğine katkı sağlamaktadır.

Araştırmadan elde edilen teorik ve uygulamalı sonuçlar doğrultusunda aşağıdaki öneriler geliştirilmiştir:

• Türk dünyasına ait ortak motifleri, süsleme sanatlarını, mimari bezemeleri, halı-kilim desenlerini ve minyatür örneklerini içeren açık erişimli, çok dilli ve semantik olarak etiketlenmiş ortak bir dijital veri tabanı oluşturulmalıdır.

• Türk Devletleri Teşkilatı, Türk Kültür ve Miras Vakfı, üniversiteler, müzeler ve kültürel miras kurumları iş birliğiyle yapay zekâ destekli kültürel arşiv ve dijital miras projeleri geliştirilmelidir.

• Yapay zekâ modellerinin eğitiminde kullanılmak üzere Türk dünyası ülkelerinin ortak katılımıyla kültürel motif veri kümeleri hazırlanmalı ve kültürel çeşitliliği temsil eden yerel veri setleri desteklenmelidir.

• Güzel sanatlar ve grafik tasarım programlarında kültürel miras çalışmaları ile yapay zekâ ve prompt mühendisliği uygulamalarını bir araya getiren disiplinlerarası eğitim modelleri geliştirilmelidir.

• Yapay zekâ ile üretilen kültürel içeriklerin etik kullanımı, fikri mülkiyet hakları, kültürel özgünlük ve dijital mirasın korunmasına yönelik akademik ve hukuki çerçeveler oluşturulmalıdır.

• Gelecekte gerçekleştirilecek araştırmalarda yalnızca amblem ve grafik tasarım uygulamaları değil; artırılmış gerçeklik (AR), sanal gerçeklik (VR), dijital oyun tasarımı, üç boyutlu modelleme, hareketli grafikler ve sanal müzecilik uygulamaları gibi farklı dijital tasarım alanları da incelenmelidir.

Sonuç olarak yapay zekâ teknolojileri, Türk dünyasının ortak görsel kültür mirasının dijital çağda korunması, yeniden yorumlanması ve gelecek kuşaklara aktarılması açısından önemli olanaklar sunmaktadır. Kültürel mirasın yalnızca müzelerde veya fiziksel eserlerde korunması değil, çağdaş dijital üretim süreçlerine entegre edilerek yaşayan ve dönüşen bir kültürel yapı hâline getirilmesi gerekmektedir. Geleneksel motiflerin yapay zekâ destekli tasarım süreçleri aracılığıyla yeniden üretilmesi, kültürel belleğin sürekliliğini destekleyen yaratıcı, sürdürülebilir ve yenilikçi bir yaklaşım olarak değerlendirilebilir.

REFERENCES

[1] UNESCO Türkiye Millî Komisyonu. (t.y.). Dünya Belleği'nde Türkiye Çalıştayı Dünya Belleği Türkiye Ulusal Listesi. https://www.unesco.org.tr/Pages/490/106/UNESCO%20T%C3%BCrkiye%20Mill%C3%AE%20Komisyonu%20D%C3%BCnya%20Belle%C4%9Fi%E2%80%99nde%20T%C3%BCrkiye%20%C3%87al%C4%B1%C5%9Ftay%C4%B1%20D%C3%BCnya%20Belle%C4%9Fi%20T%C3%BCrkiye%20Ulusal%20Listesi

[2] Stuart Hall. (1997). Representation: Cultural representations and signifying practices. Sage Publications.

[3] Jan Assmann. (2011). "Cultural memory and early civilization: Writing, remembrance, and political imagination". Cambridge University Press.

[4] Anadolu Ajansı. (2024, 13 Haziran). Türk Kültür ve Miras Vakfı, Türk dünyası kültürel mirasının dijital kataloğunu hazırlıyor. https://www.aa.com.tr/tr/kultur/turk-kultur-ve-miras-vakfi-turk-dunyasi-kulturel-mirasinin-dijital-katalogunu-hazirliyor/3248577

[5] Qazinform. (2024, 15 Nisan). Turk Heritage platform to open at Alem.ai International AI Center in Kazakh capital. https://qazinform.com/news/turk-heritage-platform-to-open-at-alemai-international-ai-center-in-kazakh-capital-187a1d

[6] Kafesoğlu, İ. (2015). Türk milli kültürü. Ötüken Neşriyat.

[7] Güngör, E. (2017). Kültür değişmesi ve milliyetçilik (15. baskı). Ötüken Neşriyat.

[8] Purtaş, F. (2013). Türk dış politikasının yükselen değeri: Kültürel diplomasi. Gazi Akademik Bakış, 7(13), 1-14. http://dergipark.org.tr/gav/issue/6515/86361

[9] Gömeç, S. Y. (2024). Türk kültürünün ana hatları. 6. Baskı. Berikan Yayınevi.

[10] Aksan, D. (2007). Her yönüyle dil: Ana çizgileriyle dilbilim (Cilt 1). Türk Dil Kurumu Yayınları.

[11] Çoruhlu, Y. (2019). Türk mitolojisinin ana hatları. Kabalcı Yayıncılık.

[12] Barnard, M. (2002). Sanat, tasarım ve görsel kültür (G. Korkmaz, Çev.). Ütopya Yayınevi.

[13] Berger, J. (2018). Görme biçimleri (Y. Salman, Çev.). Metis Yayınları. (Orijinal eser 1972 yılında yayımlanmıştır).

[14] Mirzoeff, N. (1999). An introduction to visual culture. Routledge.

[15] Felshin, N. (1995). But is it art? The spirit of art as activism. Bay Press.

[16] Ersoy, N. (2007). Semboller ve yorumları. Dönence Yayınları.

[17] Jung, C. G. (2015). İnsan ve sembolleri (A. N. Babaoğlu, Çev.). Okuyan Us Yayınları.

[18] Çoruhlu, Y. (2019). Türk mitolojisinin ana hatları. Kabalcı Yayıncılık.

[19] Barthes, R. (2016). Göstergebilimsel serüven (M. Rifat & S. Rifat, Çev.). Yapı Kredi Yayınları.

[20] Şen, S., & Bilis, A. E. (2023). Dijital kültür çağında teknoparkların yeri: Çanakkale Teknopark firma çalışanlarının görüşleri üzerine nitel bir araştırma. Uluslararası Anadolu Sosyal Bilimler Dergisi, 7(3), 578-599. https://doi.org/10.47525/ulasbid.1306608

[21] Çakır, A., & Dilci, T. (2024). Dijitalleşme, kültür ve psikoloji üçgeninde değer algısı. Cihannüma Sosyal Bilimler Akademi Dergisi, 2(2), 40-63. https://doi.org/10.55205/jocsosa.2220231461839

[22] Haines, R., & Lötter, C. (2022). Bridging the divide in cultural capital? A tale of two capitals. KOERS — Bulletin for Christian Scholarship, 87(1), 1-15. https://doi.org/10.19108/KOERS.87.1.2534

[23] Kunt, S. (2023). Dijitalleşme çağında kültürel mirası yeniden yorumlamak. Kültürel Miras Araştırmaları Dergisi, 4(2), 49-60. https://doi.org/10.59127/kulmira.1364908

[24] Kunt, S. (2023). Dijitalleşme çağında kültürel mirası yeniden yorumlamak. Kültürel Miras Araştırmaları Dergisi, 4(2), 49-60. https://doi.org/10.59127/kulmira.1364908

[25] Şen, S., & Bilis, A. E. (2023). Dijital kültür çağında teknoparkların yeri: Çanakkale Teknopark firma çalışanlarının görüşleri üzerine nitel bir araştırma. Uluslararası Anadolu Sosyal Bilimler Dergisi, 7(3), 578-599. https://doi.org/10.47525/ulasbid.1306608

[26] Çakır, A., & Dilci, T. (2024). Dijitalleşme, kültür ve psikoloji üçgeninde değer algısı. Cihannüma Sosyal Bilimler Akademi Dergisi, 2(2), 40-63. https://doi.org/10.55205/jocsosa.2220231461839

[27] Schamel, C. (2025). Frameworks, interrogations, and conceptual terrain of artificial intelligence in relation to art and culture: A content analysis. Journal of Cultural Management and Cultural Policy, 11(2), 279–291.

[28] Chen, Y. (2024). AI art and human creativity: An interactive gallery for a DAO. Advances in Education, Humanities and Social Science Research, 9, 152–165.

[29] Chen, Y. (2024). AI art and human creativity: An interactive gallery for a DAO. Advances in Education, Humanities and Social Science Research, 9, 152–165.

[30] Meyl, V. (2024). Artificial intelligence and the metamorphosis of beauty: A philosophical inquiry. Open Journal of Philosophy, 14, 180–200. https://doi.org/10.4236/ojpp.2024.141015

[31] Nurveda, A., Nursalim, M., & Masitoh, S. (2023). Aesthetics and artificial intelligence: Impact and criticism of art. Education Achievement: Journal of Science and Research, 4(3), 116–123.

[32] Fernández, D. M. M. (2024). Prospective research in the field of teaching creative skills to artificial intelligence. Interdisciplinary Cultural and Humanities Review, 3(1), 34–45. https://doi.org/10.59214/cultural/1.2024.34

[33] Nurveda, A., Nursalim, M., & Masitoh, S. (2023). Aesthetics and artificial intelligence: Impact and criticism of art. Education Achievement: Journal of Science and Research, 4(3), 116–123.

[34] Barón-Birchenall, L., Sánchez-Vallejo, A., Toro-Silva, C., & Folleco-Eraso, A. (2025). Creativity: An individual or collective phenomenon? A historical-psychological perspective. Adaptive Behavior, 33(4), 249–269.

[35] Barón-Birchenall, L., Sánchez-Vallejo, A., Toro-Silva, C., & Folleco-Eraso, A. (2025). Creativity: An individual or collective phenomenon? A historical-psychological perspective. Adaptive Behavior, 33(4), 249–269.

[36] Yazid, S. R., Mustofa, M., & Fitriyah, U. (2024). Can automatic poetry generation infuse values? Unveiling insights through content analysis of generated poetry. LiNGUA, 19(1), 70–82.

[37] Polat, B. (2023). Artificial intelligence as subject of art and democratic possibilities of art publicity. Journal of History School, 16(66), 3107–3128. https://doi.org/10.29228/Joh.72399

[38] Rane, N. L., & Choudhary, S. (2024). Role and challenges of ChatGPT, Google Bard, and similar generative artificial intelligence in arts and humanities. Studies in Humanities and Education, 5(1), 1–11. https://doi.org/10.48185/she.v5i1.999

[39] Meyl, V. (2024). Artificial intelligence and the metamorphosis of beauty: A philosophical inquiry. Open Journal of Philosophy, 14, 180–200. https://doi.org/10.4236/ojpp.2024.141015

[40] Spratt, E. L., & Elgammal, A. (2015). Computational beauty: Aesthetic judgment at the intersection of art and science. In L. Agapito, M. M. Bronstein, & C. Rother (Eds.), Computer Vision – ECCV 2014 Workshops (pp. 35–53). Springer. https://doi.org/10.1007/978-3-319-16178-5_3

[41] Meyl, V. (2024). Artificial intelligence and the metamorphosis of beauty: A philosophical inquiry. Open Journal of Philosophy, 14, 180–200. https://doi.org/10.4236/ojpp.2024.141015

[42] Arslan Kalay, H. (2022). Hayvan sembolizminin Anadolu Türk kilimlerine yansımaları. Van Yüzüncü Yıl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, (55), 40-55.

[43] Kaya, Ö., & Romanescu, S. (2021). An analogical approach to colors and symbolism in Romanian and Turkish folk art. Art-Sanat, 15, 161–181. https://doi.org/10.26650/artsanat.2021.15.0007

[44] Selçuk, K., & Yurttaş, H. (2021). Doğu Karadeniz bölgesi kirkitli dokumalarında geometrik ve bitkisel motiflerden örnekler. Art-Sanat, 15, 287-315.

[45] Berkli, Y., & Gültepe, G. (2020). Eski Türk inançlarında tabiat kültü: Erzurum Kazancı örneği. TAED – Türkiye Araştırmaları Literatür Dergisi, (67), 444-454.

[46] Biliaieva, S., Karasevych, A., & Kutsenko, S. (t.y.). Seljuk culture in Ukraine... historical and archaeological research. Sanat Tarihi Dergisi / Journal of Art History, 239-242.

[47] Kaynakcı Elinç, Z., & Kaya, L. G. (2018). Mitoloji ve inanışlar ışığında Türk kültüründe hurma ağacı (Phoenix dactylifera L.) ve sembolleri. Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 5(2), 413-424.

[48] Özgeriş, M. M., Karahan, F., & Özgeriş, M. (2022). Servi ağacının Türk kültürü ve peyzajında kazandığı anlamların Hayâlî'nin "serv" redifli kasidesindeki yansımaları. Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi, 10(1), 273-287.

[49] Kaynakcı Elinç, Z., & Kaya, L. G. (2018). Mitoloji ve inanışlar ışığında Türk kültüründe hurma ağacı (Phoenix dactylifera L.) ve sembolleri. Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 5(2), 413-424.

[50] Eroğlu, M. A., & Akyüz, T. (2022). Anadolu dokumalarında ve Kuzey Amerika Kızılderili kıyafetlerinde yıldız motifi. Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi (BAUNSOBED), 25(48), 403-420.

[51] Kayın, A., & Kalay, İ. S. (2023). Selçuklu dönemine ait bir hamam ritüeli malzemesi olarak seramik topuk törpüleri. SEFAD – Selçuk Üniversitesi Edebiyat Fakültesi Dergisi, (50), 77-100.

[52] Eroğlu, M. A., & Akyüz, T. (2022). Anadolu dokumalarında ve Kuzey Amerika Kızılderili kıyafetlerinde yıldız motifi. Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi (BAUNSOBED), 25(48), 403-420.

[53] Selçuk, K., & Yurttaş, H. (2021). Doğu Karadeniz bölgesi kirkitli dokumalarında geometrik ve bitkisel motiflerden örnekler. Art-Sanat, 15, 287-315.

[54] Duggan, T. M. P. (2019). Some notes relating to the modern misnaming of a medieval Islamic design / Ortaçağ İslam dünyasına ait bir motifin modern dönemde yanlış adlandırılmasına ilişkin notlar. MESOS, 1, 73-101.

[55] Arslan Kalay, H. (2022). Hayvan sembolizminin Anadolu Türk kilimlerine yansımaları. Van Yüzüncü Yıl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, (55), 40-55.

[56] Mohebi, M., Boroomand, B., Jalali, M., Zareian, A., Daeijavad, M., & Krunz, M. (2022). Generative adversarial networks: A game-theoretic perspective. arXiv preprint, 1-36.

[57] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27, 2672-2680.

[58] Ran, S. (2023). Applications and challenges of GAN in AI-powered artistry. Proceedings of the 2023 International Conference on Machine Learning and Automation, 41, 75-79.

[59] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27, 2672-2680.

[60] Purwono, Wulandari, A. N. E., Ma'arif, A., & Salah, W. A. (2025). Understanding generative adversarial networks (GANs): A review. Control Systems and Optimization Letters, 3(1), 36-45.

[61] Si, H. (2023). On the convergence acceleration of vanilla generative adversarial network. Journal of Physics: Conference Series, 2634(1), 1-7.

[62] Pawar, D., & Yannawar, P. (2024). Advancements and applications of generative adversarial networks: A comprehensive review. International Journal for Research in Applied Science and Engineering Technology, 12(5), 2692-2699

[63] Dash, A., Ye, J., & Wang, G. (2023). A review of Generative Adversarial Networks (GANs) and its applications in a wide variety of disciplines: From medical to remote sensing. IEEE Access, 11, 18330–18357.

[64] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

[65] Pawar, D., & Yannawar, P. (2024). Advancements and applications of generative adversarial networks: A comprehensive review. International Journal for Research in Applied Science and Engineering Technology, 12(5), 2692-2699

[66] Si, H. (2023). On the convergence acceleration of vanilla generative adversarial network. Journal of Physics: Conference Series, 2634(1), 1-7.

[67] Dash, A., Ye, J., & Wang, G. (2023). A review of Generative Adversarial Networks (GANs) and its applications in a wide variety of disciplines: From medical to remote sensing. IEEE Access, 11, 18330–18357.

[68] Dash, A., Ye, J., & Wang, G. (2023). A review of Generative Adversarial Networks (GANs) and its applications in a wide variety of disciplines: From medical to remote sensing. IEEE Access, 11, 18330–18357.

[69] Pawar, D., & Yannawar, P. (2024). Advancements and applications of generative adversarial networks: A comprehensive review. International Journal for Research in Applied Science and Engineering Technology, 12(5), 2692-2699

[70] Dash, A., Ye, J., & Wang, G. (2023). A review of Generative Adversarial Networks (GANs) and its applications in a wide variety of disciplines: From medical to remote sensing. IEEE Access, 11, 18330–18357.

[71] Purwono, Wulandari, A. N. E., Ma'arif, A., & Salah, W. A. (2025). Understanding generative adversarial networks (GANs): A review. Control Systems and Optimization Letters, 3(1), 36-45.

[72] Purwono, Wulandari, A. N. E., Ma'arif, A., & Salah, W. A. (2025). Understanding generative adversarial networks (GANs): A review. Control Systems and Optimization Letters, 3(1), 36-45.

[73] Pawar, D., & Yannawar, P. (2024). Advancements and applications of generative adversarial networks: A comprehensive review. International Journal for Research in Applied Science and Engineering Technology, 12(5), 2692-2699

[74] Dash, A., Ye, J., & Wang, G. (2023). A review of Generative Adversarial Networks (GANs) and its applications in a wide variety of disciplines: From medical to remote sensing. IEEE Access, 11, 18330–18357.

[75] Kumari, S., & Aggarwal, K. (2022). Scope of generative adversarial networks (GANs) in image processing: A review. International Journal of Health Sciences, 6(S6), 724–733.

[76] [74] Dash, A., Ye, J., & Wang, G. (2023). A review of Generative Adversarial Networks (GANs) and its applications in a wide variety of disciplines

[77] Kumari, S., & Aggarwal, K. (2022). Scope of generative adversarial networks (GANs) in image processing: A review. International Journal of Health Sciences, 6(S6), 724–733.

[78] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840-6851.

[79] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10684-10695.

[80] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10684-10695.

[81] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. arXiv preprint arXiv:2103.00020, 1-37.

[82] Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with CLIP latents. arXiv preprint arXiv:2204.06125, 1-27.

[83] Ran, S. (2023). Applications and challenges of GAN in AI-powered artistry. Proceedings of the 2023 International Conference on Machine Learning and Automation, 41, 75-79.

[84] Dash, A., Ye, J., & Wang, G. (2023). A review of Generative Adversarial Networks (GANs) and its applications in a wide variety of disciplines: From medical to remote sensing. IEEE Access, 11, 18330–18357.

[85] Rane, N. L., & Choudhary, S. (2024). Role and challenges of ChatGPT, Google Bard, and similar generative artificial intelligence in arts and humanities. Studies in Humanities a*nd Education, 5(1), 1–11. /https://doi.org/10.48185/she.v5i1.999

[86] Zailuddin, M. F. N. O., Harun, N. A. N., Rahim, H. A. A., Kamaruzaman, A. F., Berahim, M. H., Harun, M. H., & Ibrahim, Y. (2024). Redefining creative education: A case study analysis of AI in design courses. Journal of Research in Innovative Teaching & Learning, 17(2), 282–296. https://doi.org/10.1108/JRIT-01-2024-0019

[87] Osadcha, K., & Osadcha, M. (2023). Generative artificial intelligence vs humans in the process of creating corporate identity elements. Information Technologies and Learning Tools, 98(6), 213–230. https://doi.org/10.33407/itlt.v98i6.5494

[88] Chen, Y. (2024). AI art and human creativity: An interactive gallery for a DAO. Advances in Education, Humanities and Social Science Research, 9, 152–165.

[89] Zailuddin, M. F. N. O., Harun, N. A. N., Rahim, H. A. A., Kamaruzaman, A. F., Berahim, M. H., Harun, M. H., & Ibrahim, Y. (2024). Redefining creative education: A case study analysis of AI in design courses. Journal of Research in Innovative Teaching & Learning, 17(2), 282–296. https://doi.org/10.1108/JRIT-01-2024-0019

[90] Fernández, D. M. M. (2024). Prospective research in the field of teaching creative skills to artificial intelligence. *Interdisciplinary Cultural and Humanities Review*, *3*(1), 34–45. https://doi.org/10.59214/cultural/1.2024.34

[91] Fernández, D. M. M. (2024). Prospective research in the field of teaching creative skills to artificial intelligence. *Interdisciplinary Cultural and Humanities Review*, *3*(1), 34–45. https://doi.org/10.59214/cultural/1.2024.34

[92] Chen, Y. (2024). AI art and human creativity: An interactive gallery for a DAO. Advances in Education, Humanities and Social Science Research, 9, 152–165.

[93] Fernández, D. M. M. (2024). Prospective research in the field of teaching creative skills to artificial intelligence. Interdisciplinary Cultural and Humanities Review, 3(1), 34–45. https://doi.org/10.59214/cultural/1.2024.34

[94] Yazid, S. R., Mustofa, M., & Fitriyah, U. (2024). Can automatic poetry generation infuse values? Unveiling insights through content analysis of generated poetry. LiNGUA, 19(1), 70–82.

[95] Spratt, E. L., & Elgammal, A. (2015). Computational beauty: Aesthetic judgment at the intersection of art and science. In L. Agapito, M. M. Bronstein, & C. Rother (Eds.), Computer Vision – ECCV 2014 Workshops (pp. 35–53). Springer. https://doi.org/10.1007/978-3-319-16178-5_3

[96] Fernández, D. M. M. (2024). Prospective research in the field of teaching creative skills to artificial intelligence. *Interdisciplinary*

Cultural and Humanities Review, *3*(1), 34–45. https://doi.org/10.59214/cultural/1.2024.34

[97] Polat, B. (2023). Artificial intelligence as subject of art and democratic possibilities of art publicity. Journal of History School, 16(66), 3107–3128. https://doi.org/10.29228/Joh.72399

[98] Zailuddin, M. F. N. O., Harun, N. A. N., Rahim, H. A. A., Kamaruzaman, A. F., Berahim, M. H., Harun, M. H., & Ibrahim, Y. (2024). Redefining creative education: A case study analysis of AI in design courses. Journal of Research in Innovative Teaching & Learning, 17(2), 282–296. https://doi.org/10.1108/JRIT-01-2024-0019

[99] Meyl, V. (2024). Artificial intelligence and the metamorphosis of beauty: A philosophical inquiry. Open Journal of Philosophy, 14, 180–200. https://doi.org/10.4236/ojpp.2024.141015

[100] Chen, Y. (2024). AI art and human creativity: An interactive gallery for a DAO. Advances in Education, Humanities and Social Science Research, 9, 152–165.

[101] Simões, J. M., & Caldeira, W. (2024). Ethics concerns in the use of computer-generated images for human. Journal of Ethics in Higher Education, 4, 171–192.

[102] Rane, N. L., & Choudhary, S. (2024). Role and challenges of ChatGPT, Google Bard, and similar generative artificial intelligence in arts and humanities. Studies in Humanities and Education, 5(1), 1–11. https://doi.org/10.48185/she.v5i1.999

[103] He, F. (2024). Research on the application of artificial intelligence aesthetics in the cultivation of aesthetic literacy of art-normal students. International Journal of Social Sciences and Public Administration, 2(1), 160–168. https://doi.org/10.62051/ijsspa.v2n1.21

[104] Stark, L., & Crawford, K. (2019). The work of art in the age of artificial intelligence: What artists can teach us about the ethics of data practice. Surveillance & Society, 17(3/4), 442–455.

[105] Kunt, S. (2023). Dijitalleşme çağında kültürel mirası yeniden yorumlamak. Kültürel Miras Araştırmaları Dergisi, 4(2), 49-60. https://doi.org/10.59127/kulmira.1364908

[106] Kunt, S. (2023). Dijitalleşme çağında kültürel mirası yeniden yorumlamak. Kültürel Miras Araştırmaları Dergisi, 4(2), 49-60. https://doi.org/10.59127/kulmira.1364908

[107] Halaç, H. H., & Öğülmüş, V. (2021). Kültürel miras verilerinin dijital olarak depolanması: Openheritage3D örneği. The Turkish Online Journal of Design, Art and Communication, 11(2), 521-540.

[108] Zhao, H., Chen, H., Jia, G., Zheng, Q., & Wang, S. (2019). 基于字典学习的民族文化图案语义标注. Scientia Sinica Informationis, 49(2), 172‑187.

[109] Dişli, M., & Tonta, Y. (2023). Veri olarak kültürel miras koleksiyonları. Türk Kütüphaneciliği, 37(3), 191-214. https://doi.org/10.24146/tk.1317445

[110] Çakmak, T., & Yılmaz, B. (2012). Türkiye'de kültürel bellek kurumlarındaki dijitalleştirme çalışmalarının düşünce özgürlüğü bağlamında değerlendirilmesi. Bilgi Dünyası, 13(2), 418-436.

[111] Orta, N. (2024). Somut olmayan kültürel mirasın korunmasında yapay zekâ, veri bilimi ve makine öğrenmesinden yararlanma. RumeliDE Dil ve Edebiyat Araştırmaları Dergisi, 38, 748-777. https://doi.org/10.29000/rumelide.1439731

[112] Zhao, H., Chen, H., Jia, G., Zheng, Q., & Wang, S. (2019). 基于字典学习的民族文化图案语义标注. Scientia Sinica Informationis, 49(2), 172‑187.

[113] Orta, N. (2024). Somut olmayan kültürel mirasın korunmasında yapay zekâ, veri bilimi ve makine öğrenmesinden yararlanma. RumeliDE Dil ve Edebiyat Araştırmaları Dergisi, 38, 748-777. https://doi.org/10.29000/rumelide.1439731

[114] Kunt, S. (2023). Dijitalleşme çağında kültürel mirası yeniden yorumlamak. Kültürel Miras Araştırmaları Dergisi, 4(2), 49-60. https://doi.org/10.59127/kulmira.1364908

[115] Dişli, M., & Tonta, Y. (2023). Veri olarak kültürel miras koleksiyonları. Türk Kütüphaneciliği, 37(3), 191-214. https://doi.org/10.24146/tk.1317445

[116] Halaç, H. H., & Öğülmüş, V. (2021). Kültürel miras verilerinin dijital olarak depolanması: Openheritage3D örneği. The Turkish Online Journal of Design, Art and Communication, 11(2), 521-540.

[117] Halaç, H. H., & Öğülmüş, V. (2021). Kültürel miras verilerinin dijital olarak depolanması: Openheritage3D örneği. The Turkish Online Journal of Design, Art and Communication, 11(2), 521-540.

[118] Orta, N. (2024). Somut olmayan kültürel mirasın korunmasında yapay zekâ, veri bilimi ve makine öğrenmesinden yararlanma. RumeliDE Dil ve Edebiyat Araştırmaları Dergisi, 38, 748-777. https://doi.org/10.29000/rumelide.1439731

[119] Halaç, H. H., & Öğülmüş, V. (2021). Kültürel miras verilerinin dijital olarak depolanması: Openheritage3D örneği. The Turkish Online Journal of Design, Art and Communication, 11(2), 521-540.

[120] Kunt, S. (2023). Dijitalleşme çağında kültürel mirası yeniden yorumlamak. Kültürel Miras Araştırmaları Dergisi, 4(2), 49-60. https://doi.org/10.59127/kulmira.1364908

[121] Çakmak, T., & Yılmaz, B. (2012). Türkiye'de kültürel bellek kurumlarındaki dijitalleştirme çalışmalarının düşünce özgürlüğü bağlamında değerlendirilmesi. Bilgi Dünyası, 13(2), 418-436.

[122] Çakır, A., & Dilci, T. (2024). Dijitalleşme, kültür ve psikoloji üçgeninde değer algısı. Cihannüma Sosyal Bilimler Akademi Dergisi, 2(2), 40-63. https://doi.org/10.55205/jocsosa.2220231461839

[123]

İş İlişkisi Bağlamında Yapay Zekânın Etik Kullanımı

Prof. Dr. İştar Urhanoğlu
İş ve Sosyal Güvenlik Hukuku ABD.
Hacettepe Üniversitesi Hukuk Fakültesi
Ankara, Türkiye
istarurhanoglu@hacettepe.edu.tr

Doç. Dr. A. Eda Manav Özdemir
Hâkim
Ankara, Türkiye
eda_manav@hotmail.com

Arş. Gör. Damla Kaynar Kıroğlu
İş ve Sosyal Güvenlik Hukuku ABD.
Hacettepe Üniversitesi Hukuk Fakültesi
Ankara, Türkiye
damlakaynar@hacettepe.edu.tr

Arş. Gör. Beyza Aydın
İş ve Sosyal Güvenlik Hukuku ABD.
İzmir Kâtip Çelebi Üniversitesi Hukuk Fakültesi
İzmir, Türkiye
beyza.aydin@ikc.edu.tr

***Özet*— Yapay zekâ sistemleri, günümüzde çevre, sağlık, finans, kamu ve istihdam dâhil olmak üzere pek çok alanda yoğun şekilde kullanılmaktadır. Bu sistemler, iş ilişkisinin kurulmasından sona ermesine kadar tüm aşamalarda işverenlerin karar alma süreçlerine dâhil edilmektedir. Özellikle iş alım sürecinde ilanların hazırlanması, adaylara ulaşılması, özgeçmişlerin analiz edilmesi ve mülakatların yürütülmesi gibi aşamalarda; iş ilişkisinin devamında ise performans değerlendirmesi, çalışma sürelerinin belirlenmesi, görev dağılımı ve ücretlendirme gibi hususlarda yaygın olarak kullanılmaktadır.**

Yapay zekâ sistemleri, verimlilik artışı sağlamakta; zaman ve maliyet tasarrufu sunmakta ve karar alma süreçlerinde tutarlılığı güçlendirmektedir. Bununla birlikte, işçilerin kişisel verilerin ihlali, "black box" özelliği nedeniyle şeffaflık ve açıklanabilirlik sorunları, algoritmik ayrımcılık, dijital dışlanma ve işçilerin temel haklarının ihlali gibi ciddi riskleri de beraberinde getirmektedir. Bu riskler karşısında kabul edilen AB Yapay Zekâ Tüzüğü, risk temelli bir yaklaşım benimsemekte ve iş ilişkilerinde kullanılan yapay zekâ sistemlerini "yüksek riskli" kategorisinde değerlendirmektedir. Bu kapsamda Tüzük'te işverenlere kapsamlı yükümlülükler getirilmiş; özellikle çalışanların duygusal durumlarının biyometrik veriler aracılığıyla analizine dayalı sistemler yasaklanmıştır.

Türk hukukunda yapay zekâya ilişkin henüz doğrudan bir kanuni düzenleme bulunmamaktadır. Bununla birlikte düzenleyici çalışmalar hız kazanmış olup 2021-2025 Ulusal Yapay Zekâ Stratejisi hem küresel rekabet gücünün artırılmasını hem de yapay zekâ teknolojilerinin etik ve hukuki sınırlar içinde geliştirilmesini hedeflemektedir.

Bu çalışma, literatür taraması yöntemiyle iş ilişkilerinde yapay zekânın etik kullanımını uluslararası düzenlemeler ve farklı ülke uygulamaları ışığında incelemekte; Türk iş hukuku ile karşılaştırmalı bir değerlendirme yaparak mevcut düzenlemelerin yeterliliğini ve ortaya çıkan hukuki boşlukları analiz etmeyi amaçlamaktadır.

Anahtar Kelimeler—yapay zekâ etiği, iş ilişkisi, algoritmik ayrımcılık, AB Yapay Zekâ Tüzüğü, kişisel verilerin korunması.

GİRİŞ

Endüstri 4.0 ve dijitalleşme dalgası, küresel iş gücü piyasalarını ve yönetim bilimi pratiklerini köklü bir dönüşüme uğratmaktadır. Bu teknolojik dönüşümün en radikal biçimde hissedildiği alanlardan biri de kurumsal yapıların can damarını oluşturan insan kaynakları yönetimi ve özellikle işe alım süreçleridir. Geleneksel işe alım yöntemlerinin barındırdığı zaman maliyeti, idari yük ve insani hatalar; yerini hızla büyük veri analitiği (*big data analytics*), Doğal Dil İşleme (NLP) teknolojileri, makine öğrenimi (*machine learning*) ve yapay zekâ tabanlı otonom karar destek mekanizmalarına bırakmaktadır. Günümüzde modern işletmeler; reklamcılık ve algoritmik profilleme yoluyla pasif adaylara ulaşmaktan, otomatik CV ayrıştırma ve puanlama sistemlerine; yapay zekâ destekli video mülakat analizlerinden oyun tabanlı bilişsel testlere kadar uzanan geniş bir dijital araç setini operasyonel hale getirmiştir. Bu otonom sistemlerin temel kurumsal vaadi; insan kaynakları uzmanlarının subjektif ön yargılarından arındırılmış, veri temelli, rasyonel, standart ve hatasız bir seçme-yerleştirme mekanizması inşa ederek operasyonel verimliliği maksimize etmektir.

Ancak yapay zekânın sunduğu bu yadsınamaz hız, maliyet avantajı ve kaynak tasarrufu pratikleri, madalyonun diğer yüzünde çözülmesi gereken karmaşık hukuki, etik ve sosyoekonomik krizleri de beraberinde getirmektedir. Literatürde "kara kutu" (*black box*) problemi olarak adlandırılan algoritmik şeffaflık eksikliği; adayların neden elendiklerini denetleyemedikleri yapısal bir korumasızlık alanı yaratmaktadır[1]. Daha da önemlisi, yapay zekânın eğitildiği tarihsel veri setlerindeki gizli yanlılıklar, korumalı kategoriler olan cinsiyet, etnik köken, yaş ve engellilik temelinde yeni nesil bir "algoritmik ayrımcılık" riskini doğurmaktadır. Adayların rızası ve işin niteliğiyle illiyet bağı bulunmaksızın e-posta

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

kutularının, sosyal medya beğenilerinin, izleme ve tüketim alışkanlıklarının otonom yazılımlarla taranması ise özel hayatın gizliliği, veri minimizasyonu ve insan onuru ilkelerini derinden zedelemektedir[2]. Nitekim küresel ölçekte yaşanan *Amazon*, *Meta* ve *Workday* gibi emsal davalar, yapay zekâ sistemleri tarafından gerçekleştirilen algoritmik ayrımcılıkların somut örnekleridir[1].

Dijitalleşmenin yarattığı bu asimetrik riskler karşısında küresel hukuk düzenleri de proaktif koruma kalkanları geliştirme arayışına girmiştir. Avrupa Birliği, Genel Veri Koruma Tüzüğü'nün (GDPR) otomatik karar vermeyi sınırlandıran emredici hükümleriyle başlattığı regülasyon hamlesini; yakın zamanda yürürlüğe giren ve işe alım algoritmalarını doğrudan "yüksek riskli" kategoride konumlandıran AB Yapay Zekâ Tüzüğü ve algoritmik yönetime sınırlar çizen Platform Çalışması Direktifi ile evrensel bir statüye taşımıştır. Türk hukuku normatif ekosisteminde ise henüz müstakil bir yapay zekâ kanunu bulunmamakla birlikte; 6698 sayılı Kişisel Verilerin Korunması Kanunu'nun (KVKK) emredici ilkeleri, mülga Kişisel Verileri Koruma Kurulu'nun yayımladığı yapay zekâ kılavuzları, Ulusal Yapay Zekâ Stratejisi (2021-2025) ile 2024-2025 Eylem Planı ve nihayetinde mevzuatın AB Yapay Zekâ Yasası ile uyumlaştırılacağını ilan eden 2025-2027 Orta Vadeli Program (OVP), ulusal düzeydeki hukuki dönüşümün yönünü ve kararlılığını göstermektedir.

Bu çalışmanın temel amacı ve özgün değeri; işe alım süreçlerinde yapay zekâ kullanımının kurumsal avantajlarını ortaya koyarken, bu uygulamaların adayların kişilik hakları, veri koruma hukuku ve ayrımcılık yasağı bağlamında doğurduğu riskleri disiplinler arası bir perspektifle analiz etmektir. Çalışma kapsamında öncelikle aday tarama, özgeçmiş ayrıştırma ve mülakat otomasyonunda kullanılan güncel teknolojiler ve kurumsal vaka analizleri ele alınacak; ardından bu süreçlerin GDPR, KVKK ve AB Yapay Zekâ Tüzüğü eksenindeki yasal sınırları tartışmaya açılacaktır. Son aşamada ise, dijital emek piyasasında operasyonel verimlilik ile temel insan hakları arasında adil bir dengenin kurulabilmesi adına, insan denetimini ve şeffaflığı merkeze alan "hibrit karar modeli" önerisi temelinde hukuki ve yönetsel çözüm tavsiyeleri geliştirilecektir.

I. YAPAY ZEKÂ VE İŞE ALIM SÜREÇLERİNİN DÖNÜŞÜMÜ

Dünya Ekonomik Forumu'nun raporuna göre, 2027 yılına kadar mevcut işlerin yaklaşık %23'ünün dönüşüme uğrayacağı belirtilmektedir. Bu süreçte yaklaşık 69 milyon yeni iş alanının ortaya çıkması beklenirken, 83 milyon işin ortadan kalkacağı öngörülmekte; böylece toplamda 14 milyonluk net bir istihdam kaybı ortaya çıkmaktadır. Özellikle yapay zekâ uzmanları, veri analistleri, siber güvenlik uzmanları ve sürdürülebilirlik alanında çalışan uzmanların en hızlı büyüyen meslek grupları arasında yer alacağı ifade edilmektedir. Bu durum, dijitalleşme ve yeşil dönüşüm süreçlerinin işgücü piyasası üzerindeki etkisini açıkça ortaya koymaktadır[2].

Uluslararası Çalışma Örgütü'nün (ILO) 2025 tarihli "Generative AI Exposure Index" çalışması da benzer şekilde üretken yapay zekânın çalışma hayatı üzerindeki etkilerine dikkat çekmektedir. Rapora göre, dünya genelinde her dört işçiden biri, görevlerinin en az bir kısmı üretken yapay zekâya maruz kalan mesleklerde çalışmaktadır. Özellikle büro ve sekreterlik işleri, yapay zekâdan etkilenme düzeyi en yüksek meslek grupları arasında gösterilmektedir. Bunun yanında kadın çalışanların yapay zekâya maruziyet oranının erkeklere kıyasla daha yüksek olduğu belirtilmektedir. Bununla birlikte ILO, yapay zekânın çoğu meslek bakımından tam anlamıyla bir ikame yaratmayacağını, esas etkinin işlerin tamamen ortadan kalkmasından ziyade görevlerin ve çalışma süreçlerinin dönüşmesi şeklinde ortaya çıkacağını vurgulamaktadır. Raporda profesyonel meslekler, teknik elemanlar ile makine operatörleri ve montajcılar bakımından istihdam artışı beklendiği; buna karşılık büro hizmetleri, hizmet ve satış işleri ile nitelik gerektirmeyen mesleklerde daralma yaşanmasının öngörüldüğü ifade edilmektedir[3].

II. YAPAY ZEKANIN TANIMI VE KULLANIM ALANLARI

Yapay zekanın kesin ve herkes tarafından kabul edilen bir tanımı yoktur[4]. Türk Dil Kurumu yapay zekayı, "Bir bilgisayarın, bilgisayar kontrolündeki bir robotun veya programlanabilir bir aygıtın insana benzer biçimde algılama, öğrenme, fikir yürütme, karar verme, sorun çözme, iletişim kurma vb. işlevleri sergileyebilme yeteneği" olarak tanımlamaktadır.

Ekonomik Kalkınma ve İşbirliği Örgütü (OECD)'nün Yapay Zekâya İlişkin Tavsiye Kararı'nda (OECD 2019) yapay zekâ şöyle tanımlanmıştır: "*Bir yapay zekâ sistemi, açık veya örtük*

[1] Bkz. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G/ [11.05.2026]

[2] Bkz. https://www.weforum.org/press/2023/04/future-of-jobs-report-2023-up-to-a-quarter-of-jobs-expected-to-change-in-next-five-years/ [07.05.2026]

[3] Bkz. https://www.ilo.org/publications/generative-ai-and-jobs-refined-global-index-occupational-exposure

[4] https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/10/g7-toolkit-for-artificial-intelligence-in-the-public-sector_f93fb9fb/421c1244-en.pdf [07.05.2026]

hedefler doğrultusunda, aldığı girdilerden hareketle tahminler, içerik, öneriler veya kararlar gibi çıktıları nasıl üreteceğini çıkarım yoluyla belirleyen ve bu çıktılar aracılığıyla fiziksel ya da sanal ortamları etkileyebilen makine temelli bir sistemdir. Farklı yapay zekâ sistemleri, uygulamaya alındıktan sonra sahip oldukları özerklik ve uyarlanabilirlik düzeyleri bakımından farklılık gösterebilir[3]".

Yapay zekâ, sağlık, eğitim, emniyet, bankacılık, hizmet, eğlence ve hizmet sektörü gibi birçok farklı alanda kullanılmaktadır. Özellikle kişisel asistanlar (örneğin Siri, Google Asistan ve Amazon Alexa), sürücüsüz araçlar, sağlık hizmetleri, e-ticaret ve reklamcılık, finansal hizmetler, üretim ve lojistik ile oyun sektörü yapay zekâ kullanımının yaygın olduğu alanlar arasında yer almaktadır. Ayrıca görüntü ve ses tanıma teknolojileri de yüz tanıma, nesne tanıma, konuşma tanıma ve duygu analizi gibi uygulamalar aracılığıyla giderek daha fazla önem kazanmaktadır.

Teknik açıdan yapay zekâ sistemleri yazılım tabanlı veya donanım tabanlı olabilmektedir. Yazılım tabanlı yapay zekâya sesli asistanlar ve yüz tanıma sistemleri örnek gösterilebilirken; donanım tabanlı yapay zekâ kapsamında robotlar ve otonom araçlar öne çıkmaktadır.

III. YAPAY ZEKÂNIN İŞ HUKUKUNDA KULLANIMI

Yapay zekâ teknolojilerinin çalışma hayatında kullanımı, aday belirleme, ücret tespiti ve işe alım süreçlerinin otomatikleştirilmesini mümkün hâle getirmektedir. Bununla birlikte, özellikle duygu analizi ve "people analytics" uygulamaları, mahremiyetin ihlali ve kişisel verilerin gizliliği bakımından önemli hukuki riskler doğurmaktadır. Yapay zekâ kullanımının yaygınlaşması, iş hukuku, kişisel verilerin korunması hukuku ve ayrımcılık yasağı bakımından yoğun tartışmaları beraberinde getirmiştir. Ayrıca bu teknolojilerin çalışma ilişkilerinde kullanılması, işçilerin kişilik hakları, iş güvencesi ve sendikal hakları açısından da çeşitli riskler yaratabilmektedir[4].

A. İşe Alımda Yapay Zekâ Kullanımı

İstihdam ilişkilerinde karşılaşılan temel sorunlardan biri, "doğru işe doğru insan" ilkesinin yeterince etkin şekilde hayata geçirilememesidir. Kurumsal başarı ve sürdürülebilir değer üretimi, büyük ölçüde nitelikli insan kaynağının doğru biçimde seçilmesine bağlıdır. Geleneksel işe alım süreçlerinde karar verme mekanizması çoğu zaman işverenin sınırlı bilgiye dayalı değerlendirmelerine dayanmakta ve adayların uygun niteliklerinin belirlenmesi zaman ile maliyet gerektirmektedir. Dijitalleşme ile birlikte işletmeler, işe alım süreçlerinde giderek daha fazla yapay zekâ tabanlı araçlardan yararlanmaya başlamıştır. Yapay zekâ; özgeçmiş tarama, aday eşleştirme, profilleme ve mülakat analizi gibi alanlarda kullanılmaktadır. Bu sistemler, zaman ve maliyet tasarrufu sağlamanın yanı sıra karar alma süreçlerinin daha veri temelli ve objektif bir yapıya kavuşmasına katkı sunmaktadır.

Yapay zekâ destekli işe alım süreçleri, ilan hazırlama ve yayımlama aşamasından başlayarak adayların pasif biçimde profillenmesi ve sürece dâhil edilmesine kadar uzanan geniş bir kullanım alanına sahiptir. Ayrıca sosyal medya taramaları, adayların sıralanması ve elenmesi süreçleri, beceri ve kişilik testleri ile mülakat aşamalarında da yapay zekâ teknolojilerinden yararlanılmaktadır [5].

1. İş ilanlarının Yayınlanması ve Hazırlanması

Dijitalleşme ve yapay zekâ teknolojileri, günümüz iş gücü piyasasında insan kaynakları yönetimini köklü bir değişime uğratmaktadır. Bu kapsamda yapay zekâ; iş ilanlarının optimize edilerek hazırlanması, en uygun ilan platformlarının seçimi ve süreçlerin dijital kanallar üzerinden yayımlanması aşamalarında etkin bir rol oynamaktadır. Özellikle hedefli reklamcılık (targeted advertising) stratejileri doğrultusunda, hazırlanan iş ilanları belirli aday gruplarına algoritmik profilleme yöntemleri kullanılarak ulaştırılmaktadır. Algoritmik profilleme; adaylara ait kişisel verilerin derinlemesine analiz edilmesi suretiyle, bireylerin yetkinlikleri, eğilimleri ve işe uygunlukları hakkında öngörüsel (predictive) profiller oluşturulmasını ifade eder. Bu doğrultuda devreye giren yapay zekâ tabanlı eşleştirme sistemleri, aday-iş uyumunu dinamik bir biçimde analiz ederek geleneksel ve zaman alıcı olan işe alım süreçlerini önemli ölçüde hızlandırmaktadır[5].

Söz konusu teknolojik dönüşüm, süreçleri hızlandırmanın yanı sıra işe alımda objektifliğin sağlanması ve ayrımcılığın önlenmesi noktasında da ciddi bir potansiyel barındırmaktadır. Güvenilir, tarafsız ve nitelikli verilerle (unbiased data) eğitilen yapay zekâ sistemleri, insan kaynakları uzmanlarının karar alma süreçlerinde ortaya çıkabilecek bilinçaltı ön yargılarını (unconscious bias) asgariye indirme kapasitesine sahiptir. Bu duruma somut bir örnek olarak *Textio* uygulaması gösterilebilir. *Textio*, iş ilanlarını ve adaylarla kurulan iletişimi dilbilimsel açıdan analiz ederek metinlerdeki ayrımcı veya toplumsal cinsiyet yanlısı (gender-biased) ifadeleri tespit etmektedir. Uygulama, ilan metinlerinde başvuru oranlarını doğrudan etkileyen dil kalıplarını belirlemekte ve işverenlerin daha kapsayıcı (inclusive) bir içerik dili oluşturmasına katkı sunmaktadır. Böylelikle, anahtar kelime analizleri ve doğru iletişim kanallarının tespiti yoluyla potansiyel aday havuzuna erişim maksimum düzeye çıkarılmaktadır[5].

[5] Bkz. https://textio.com/ [12.05.2026].

Yapay zekanın nesnel bir seçme zemini yaratmasındaki bir diğer önemli teknolojik araç ise *Taledo* uygulamasıdır. *Taledo*, adayların fotoğraflarını bulanıklaştıran ve isimlerini yalnızca baş harflerine indirgeyen özel filtreler kullanarak akademik ve pratik literatürde "kör işe alım" (*blind hiring*) olarak adlandırılan yaklaşımı dijital ortamda desteklemektedir. Bu tür algoritmik filtreleme mekanizmaları sayesinde, başvuru sahiplerinin cinsiyeti, etnik kökeni, dini veya dış görünüşü gibi liyakat dışı unsurlar tamamen arka plana itilmektedir. Sonuç olarak, adayların yalnızca nesnel nitelikleri, yetkinlikleri ve mesleki becerileri temelinde değerlendirilmesi mümkün hale gelerek, fırsat eşitliğine dayalı ve adil bir işe alım prosedürünün işletilmesi kolaylaşmaktadır[6].

2. Pasif Adayları Sürece Dahil Etme

İş gücü piyasasında yapay zekâ destekli işe alım süreçlerinin sağladığı bir diğer önemli avantaj, geleneksel yöntemlerle ulaşılması güç olan "pasif adaylar" (*passive candidates*) havuzunun etkin bir biçimde sürece dahil edilebilmesidir. Literatürde pasif adaylar; aktif olarak bir iş arayışı bulunmayan, iş ilanlarına doğrudan başvuruda bulunmayan, kariyer ve iş arama sitelerindeki profillerini güncellemeyen ve çoğunlukla mevcut istihdam ilişkisinden ve çalışma şartlarından memnun olan nitelikli iş gücü olarak tanımlanmaktadır. Bu aday gruplarının potansiyel iş teklifleri için tespit edilebilmesi amacıyla; sosyal medya platformları, LinkedIn gibi profesyonel ağlar, kişisel veya mesleki bloglar ve diğer dijital çevrimiçi yayınlar üzerinden adaya dair yapılandırılmamış veriler toplanmaktadır[6].

Bu noktada yapay zekâ teknolojileri, geleneksel arama motorlarının kapasitesini aşarak arama ve tarama sonuçlarını çok daha geniş, dinamik ve ayrıntılı bir biçimde geliştirmektedir. Gelişmiş entegrasyon mekanizmaları sayesinde adaylar, özgeçmişlerini doğrudan sisteme yükleyebilmekte veya dijital ve sosyal ağlardaki güncel verilerini doğrudan kariyer portallarına bağlayabilmektedir. Yapay zekâ tabanlı algoritmalar; adayların eğitim geçmişi ve edindikleri mesleki sertifikalar dâhil olmak üzere tüm kariyer yolunu bütüncül bir yaklaşımla analiz etmekte ve iş portallarındaki devasa veri yığınları arasından en uygun sonuçları süzerek filtrelemektedir. Söz konusu akıllı filtreleme ve değerlendirme sürecinde, adayın biçimsel eğitim ve mesleki niteliklerinin yanı sıra; spesifik beceri ve yetkinlikleri, mesleki olgunluk düzeyi, kıdemi, kronolojik kariyer gelişimi ve coğrafi konumu gibi çok boyutlu parametreler de eş zamanlı olarak dikkate alınmaktadır. Böylelikle yapay zekâ, işverenler için doğru yeteneğe zamanında ve nokta atışı bir biçimde ulaşılmasını sağlayan stratejik bir araç işlevi görmektedir.

3. Sosyal Medya Taramaları

Yapay zekâ destekli işe alım araçlarının pasif adaylara ulaşmada ve geniş veri yığınlarını analiz etmede sağladığı kolaylıklar, beraberinde ciddi hukuki ve etik tartışmaları da getirmektedir[6]. Bu süreçte özellikle adayların dijital ayak izlerinin takip edilmesi esasına dayanan sosyal medya taramaları, iş ilişkisinin kurulmasından önceki dürüstlük kuralı ve işverenin aydınlatma yükümlülüğü sınırlarını zorlayan riskler barındırmaktadır. Yapay zekâ tabanlı algoritmalar aracılığıyla gerçekleştirilen bu taramalarda, adayın doğrudan işe uygunluğu ile illiyet bağı bulunmayan ve esasen adayın özel hayatının gizliliğine müdahale teşkil edecek nitelikteki çeşitli verilerin, "en iyi adayın" tespit edilmesinde birer seçim kriteri olarak dikkate alınması ihtimali söz konusudur.

Dijital ekosistemde gerçekleştirilen bu izlemeler neticesinde, sosyal ağlar üzerinden adayın arkadaşlık ilişkileri, sosyal ve mesleki çevresi, dijital platformlarda ne tür paylaşımlar yaptığı ve hatta hangi içerikleri beğendiği gibi detaylı veriler elde edilebilmektedir. Bunun da ötesinde, adayın kişisel seyahat ve eğlence anlayışı ile tüketim alışkanlıklarına ilişkin profilleme yapılmasına imkân veren hassas veriler de yapay zekâ tarafından edinilebilmekte ve sistematik bir değerlendirmeye tabi tutulabilmektedir. İş ilişkisinin bağımlılık unsuru ve işçinin korunması ilkesi gözetildiğinde; adayın rızası veya işin niteliğiyle doğrudan ilgili meşru bir menfaat bulunmaksızın, özel yaşam alanına dahil olan bu verilerin işe alım kararlarına dayanak yapılması, başta kişisel verilerin korunması hakkı olmak üzere temel kişilik haklarının ihlali riskini doğurmaktadır.

Yapay zekâ algoritmalarının adayların dijital ayak izlerini analiz ederken kullandığı veri setleri; Netflix izleme geçmişinden Spotify listelerine, Facebook ve Twitter (X) gibi platformlardaki paylaşımlardan beğeni tuşlamalarına, seyahat tercihlerinden tüketim alışkanlıklarına kadar uzanan çok geniş ve heterojen bir alana yayılmaktadır. Bu veri setlerinin makine öğrenimi algoritmaları tarafından korelasyonel analizlere tabi tutulması, iş gücü piyasasında yeni nesil bir ayrımcılık ve dışlanma riskini beraberinde getirmektedir. Zira bu sistemler, adayın doğrudan mesleki yetkinliğiyle hiçbir illiyet bağı bulunmayan kişisel eğilimlerinden yola çıkarak, adayın gelecekteki iş performansına veya sadakatine yönelik spekülatif öngörülerde (predictive profiling) bulunmaktadır.

Bu durumun en somut ve tehlikeli sonucu, belirli aday gruplarının daha başvuru aşamasına bile gelemeden sistem tarafından elenmesi ya da hedefli reklam algoritmaları nedeniyle ilgili iş ilanlarını dijital ekranlarında hiç görememesidir (*algorithmic exclusion*)[7]. Örneğin, bir adayın Netflix üzerinden "çok fazla belirsiz/farklı belgesel izlemesi", makine öğrenimi algoritması tarafından *"bu işi oldukça sıkıcı bulacağı ve en kısa sürede başka bir iş arayacağı"* şeklinde yorumlanarak adayın elenmesine gerekçe yapılabilmektedir.

[6] Bkz. https://www.taledo.com/de/ [12.05.2026]

[7] Bkz. https://www.appcast.io/pay-per-applicant-not-per-click/ [13.05.2026]

Benzer şekilde, adayın sosyal medya platformlarında *"kendisi hakkında çok fazla tweet atması veya Facebook'ta yoğun olarak selfie paylaşması"*, algoritma tarafından *"iyi bir takım oyuncusu olamayacağına"* dair kesin bir karine olarak kabul edilerek adayın mülakat aşamasına geçmesi engellenebilmektedir.

Hukuki açıdan bakıldığında bu süreç, algoritmaların "kara kutu" (*black box*) niteliği taşıması sebebiyle ciddi bir şeffaflık ve denetlenebilirlik sorunu yaratmaktadır. Adaylar, neden elendiklerini bilmedikleri gibi, sistemin arkasında çalışan veri modellerinin hangi ön yargılı parametreleri esas aldığından da habersiz kalmaktadır. İş ilişkisinin kurulmasından önceki sözleşme öncesi süreçte (culpa in contrahendo) işverenin dürüstlük kuralına uygun davranma ve eşit işlem yapma borcu dikkate alındığında; rasyonel, nesnel ve denetlenebilir olmayan bu tür algoritmik çıkarımlara dayanarak adayların elenmesi, başta ayrımcılık yasağı olmak üzere adil işe alım ilkelerinin ve temel kişilik haklarının açık bir ihlali niteliğindedir[1].

4. Aday Tarama ve Özgeçmiş Analizi

İnsan kaynakları yönetiminde yapay zekâ entegrasyonunun en yoğun görüldüğü alanlardan biri de aday tarama ve özgeçmiş analizi süreçleridir. Yapay zekâ tabanlı sistemler; internet ekosistemini ve dijital işe alım platformlarını proaktif bir biçimde tarayarak, açık pozisyonlara en uygun adayları aktif bir şekilde belirleme yetkinliğine sahiptir. Bu süreçte, adayların dijital profillerinin ve sektörel ayak izlerinin değerlendirilmesinde sosyal medya analizleri ile Doğal Dil İşleme (*Natural Language Processing - NLP*) tekniklerinden yararlanılmakta ve elde edilen veriler ışığında adaylar arasında otomatik bir sıralama (*automated ranking*) gerçekleştirilmektedir. Özellikle makine öğrenimi destekli CV tarama sistemleri, insan gücünün günler alabileceği yüzlerce özgeçmişi çok kısa bir süre zarfında analiz ederek nesnel aday-iş uyumunu (*candidate-job fit*) matematiksel modellerle ölçmektedir. Bu ekosistemde kullanılan "CV ayrıştırma" (*CV parsing*) teknolojileri ise özgeçmişlerdeki dağınık ve yapılandırılmamış verileri anlamlandırarak, kurumsal Aday Takip Sistemlerine (*Applicant Tracking Systems - ATS*) otomatik olarak aktarmakta ve veri girişindeki hata payını sıfıra indirmektedir. Gelişmiş denetimli öğrenme modelleriyle donatılan bu sistemler, kurumun geçmiş dönemlerdeki başarılı işe alım verilerinden (tarihsel veri setleri) sürekli olarak öğrenerek mevcut özgeçmişleri puanlamakta (*resume scoring*) ve insan kaynakları uzmanlarının karar alma süreçlerine stratejik bir tahmin desteği sunmaktadır[5].

Sürecin aday ile temas kurulan ön aşamalarında ise yapay zekâ destekli sohbet robotları (*conversational AI/chatbots*) başvuru verilerini doğrudan toplayarak işe alım sürecinin otomasyon derecesini en üst seviyeye çıkarmaktadır. Sektörel uygulamalar incelendiğinde bu otomasyonun farklı yapay zekâ mimarileriyle hayata geçirildiği görülmektedir. Örneğin; *Paradox* uygulaması, aday ile dinamik bir diyalog kurarak başvuruya yönelik nitelikli verileri arka planda yapılandırılmış bir biçimde toplamaktadır. Benzer şekilde *Phenom* chatbot, yapay zekâ destekli sohbet arayüzü üzerinden adaylar ile ilk mülakat simülasyonlarını gerçekleştirerek işe alım evrelerini insan müdahalesi olmaksızın otomatikleştirmektedir. Aday etkileşimini bir adım ileriye taşıyan *HireVue* diyalogsal yapay zekâ (*conversational AI*) uygulaması ise, aday ile gerçekleştirilen sesli ve yazılı görüşmeler üzerinden gelişmiş veri çıkarımları gerçekleştirmektedir. Bu noktada devreye giren "video analizi" teknolojileri, adayın yalnızca sözel ifadelerini değil; mülakat esnasındaki mimiklerini, jestlerini, ses tonu dalgalanmalarını ve mikro ifadelerini de analitik birer veriye dönüştürerek adayın yetkinlik ve kurumsal kültür uyumuna dair derinlemesine psikometrik çıktılar üretmektedir.

5. Sıralama/Eleme Süreçleri

İşe alım süreçlerinin ikinci ve en kritik aşamasını oluşturan sıralama ve eleme süreçleri, yapay zekâ optimizasyonu sayesinde insan kaynakları yönetiminde operasyonel verimliliği en üst düzeye çıkaran evrelerden biridir. Bu aşamada, özgeçmişlerde ve başvuru formlarında yer alan belirli anahtar kelimeleri, teknik terimleri ve anlamsal kalıpları tanımak üzere eğitilen gelişmiş algoritmalar vasıtasıyla adaylar; açık iş ilanının gerektirdiği spesifik beceri, yetkinlik ve deneyim kriterlerine göre hiyerarşik bir biçimde sıralanabilmektedir. Tarama sonucunda elde edilen veriler ve analitik sonuçlar eş zamanlı olarak değerlendirilmekte; adaylar sistem tarafından otomatik olarak puanlandırılarak (*automated scoring*), pozisyona en uygun profillerden oluşan bir "kısa liste" (*shortlist*) dinamik bir biçimde oluşturulmaktadır. Yapay zekânın ön eleme sürecini bu şekilde otomatikleştirmesi, geleneksel işe alım prosedürlerinde yaşanan zaman kayıplarını engelleyerek toplam işe alım süresini (*time-to-hire*) önemli ölçüde kısaltmaktadır. Sürecin hızlanmasıyla sağlanan bu kayda değer zaman tasarrufu, insan kaynakları uzmanlarının idari yükünü hafifletmekte ve listenin üst sıralarında yer alan nitelikli adayların mülakat aşamalarında daha kapsamlı, derinlemesine ve nitelikli bir değerlendirmeye tabi tutulmasına imkân tanımaktadır.

Ancak, sıralama ve eleme süreçlerinin tamamen algoritmalara devredilmesi, iş hukuku ve aday hakları açısından "sistematik elenme" riskini de beraberinde getirmektedir. Yapay zekâ tabanlı algoritmik filtreleme mekanizmaları, belirli veri modellerini ve katı kalıpları esas aldığından; adayın özgeçmişindeki biçimsel bir farklılık, kariyer boşluğu veya algoritmanın ön yargıyla kodladığı bir parametre, adayın sistem tarafından doğrudan dışlanmasına (*algorithmic exclusion*) yol açabilmektedir. Bu durumun en tehlikeli boyutu, benzer yapay zekâ altyapılarını ve benzer veri setlerini kullanan farklı işverenler bakımından da adayın sistematik ve tekrar eden bir şekilde elenmesine sebebiyet vermesidir. Adayın, bir algoritma hatası veya yanlılığı nedeniyle iş gücü piyasasındaki pek çok iş ilanında eş zamanlı olarak görünmez hale gelmesi ya da otomatik olarak elenmesi, bireyin çalışma hakkını ve sözleşme

özgürlüğünü dolaylı olarak kısıtlayan, hukuki denetimi son derece zor yeni bir yapısal bariyer yaratmaktadır[5].

6. Mülakat Süreçleri

İşe alım süreçlerinin en dinamik ve insan odaklı evresi olan mülakatlar, günümüzde yapay zekânın otomasyon ve derin analitik kabiliyetleri sayesinde yeni bir boyut kazanmıştır. Yapay zekâ tabanlı sistemler, mülakatların zamanlanması ve lojistik olarak planlanması gibi idari süreçleri tamamen otomatikleştirebilmenin ötesinde, mülakatların bizzat yürütülmesinde de aktif birer aktör haline gelmiştir. Özellikle asenkron video mülakatlar ile telefon görüşmelerinde devreye giren algoritmalar; adayların ses tonunu, yüz ifadelerini, mikro mimiklerini ve genel iletişim becerilerini anlık olarak incelemektedir. Bu süreçte ses tonlarındaki en ufak frekans değişiklikleri, duraklamalar, vurgular ve nüanslar ile adayın kelime seçimleri ve dil kullanımı (sentaks ve semantik özellikleri) gelişmiş algoritmalar tarafından sistematik olarak analiz edilmektedir. Metin ve konuşma analizi yöntemleri (Doğal Dil İşleme - NLP) kullanılarak gerçekleştirilen bu dijital incelemeler, adayın yalnızca yüzeysel cevaplarını değil; kişilik özelliklerini, psikolojik profillerini ve geleceğe yönelik davranışsal eğilimlerini de haritalandırmaktadır. Böylelikle, oldukça kısa süreli mülakat verilerinden dahi adayın iletişim becerisi, içsel motivasyonu, stres altındaki direnci ve anlık duygusal durumu (*emotional state*) hakkında kapsamlı değerlendirmeler yapılabilmektedir.

Sektörde bu teknolojiyi operasyonel hale getiren kurumsal uygulamalar, yapay zekânın ulaştığı analiz kapasitesini somut bir biçimde ortaya koymaktadır. Bu doğrultuda, Almanya merkezli *Precire Technologies GmbH* tarafından geliştirilen *Precire* yazılımı, dikkat çekici bir örnek teşkil etmektedir. Şirketin verilerine göre söz konusu yazılım, başvuru sahibi ile gerçekleştirilen yalnızca 15 dakikalık kısa bir ses/konuşma mülakatının ardından, adayın son derece doğru, tutarlı ve kapsamlı bir kişilik profilini (*personality profiling*) otonom olarak oluşturabilmektedir. Benzer şekilde, küresel ölçekte faaliyet gösteren Amerikan şirketi *HireVue*, gelişmiş konuşma analizi teknolojisini derin video işleme fonksiyonlarıyla birleştirmiştir. Bu entegrasyon sayesinde sistem, mülakat esnasında adayın yüzünde beliren ve insan gözünün ayırt etmekte zorlanacağı anlık mikro ifadeleri (*micro-expressions*) analiz edebilmekte; böylelikle başvuru sahibinin o anki ruh hali, dürüstlük düzeyi ve kurumsal aidiyet potansiyeli hakkında analitik sonuçlara varabilmektedir.

Aday verilerinin işlenmesini mülakat odasının dışına taşıyarak süreci daha da radikalleştiren bir diğer girişim ise Finlandiya menşeili *DigitalMinds*'tır. Bu girişim; adayların özel e-posta kutularını ve kişisel/profesyonel sosyal ağ hesaplarını yapay zekâ algoritmalarıyla otomatik olarak analiz etmeyi, böylece adayların başvuru formları doldurma yükümlülüğünü ortadan kaldırarak işe alım sürecini tamamen basitleştirmeyi amaçlamaktadır. Ancak, *Precire* ve *HireVue*'nun sunduğu psikometrik/biyometrik analizler ile *DigitalMinds*'ın özel iletişim kanallarına yönelik bu derin tarama faaliyetleri, iş hukuku ve kişilik hakları bağlamında ciddi riskleri de beraberinde getirmektedir. Adayın ruh halinin, ses tonunun veya dijital yazışmalarının bir makine tarafından "puanlanması", işe alımda şeffaflık ilkesini zedelediği gibi; adayın rızasının sınırlarını, insan onurunun korunmasını ve özel hayatın gizliliğine yönelik orantısız müdahaleleri tartışmaya açan kritik bir hukuki gri alan yaratmaktadır.

7. Beceri/Kişilik Testleri

Yapay zekâ teknolojilerinin işe alım süreçlerine entegrasyonu, geleneksel ve statik nitelikteki çoktan seçmeli psikometrik testlerin yerini dinamik, etkileşimli ve otonom ölçme-değerlendirme metodolojilerinin almasına yol açmıştır. Bu dönüşümün en belirgin görüldüğü alanlardan biri, adayların bilişsel yeteneklerini ve işle ilgili spesifik becerilerini ölçümlemeyi hedefleyen yeni nesil beceri ve kişilik testleridir. Günümüzde yapay zekâ destekli sistemler; özellikle "oyun tabanlı testler" (*game-based assessments*) ve senaryo temelli dijital simülasyonlar aracılığıyla adayların problem çözme hızı, kriz anında karar verme mekanizmaları, analitik yetkinlikleri ve stratejik düşünme gibi üst düzey bilişsel becerilerini eş zamanlı olarak değerlendirebilmektedir. Oyunlaştırma (*gamification*) yöntemleriyle birleştirilen bu algoritmik testler, adayın yapay bir test ortamındaki stresini azaltmayı amaçlarken, arka planda adayın her bir hamlesini, harcadığı süreyi ve risk alma eğilimini veri kümesine dönüştürerek analiz etmektedir[8].

Bu tür dijital süreçler, adayın mesleki becerilerini ölçmenin yanı sıra, doğrudan adayın karakter yapısını ve kurumsal kültüre uyumunu ortaya koyan derinlemesine kişilik analizlerinin (*personality profiling*) çıkarılmasında da aktif olarak kullanılmaktadır. Yapay zekâ, adayın simülasyonlar esnasındaki davranışsal reflekslerinden yola çıkarak, beş faktörlü kişilik modeli (*Big Five*) gibi bilimsel ölçeklerle uyumlu otonom psikometrik çıktılar üretmektedir. Ancak bu otonom test ve analiz süreçleri, iş hukuku doktrininde adil yargılanma ve şeffaflık ilkeleri bağlamında ciddi tartışmaları beraberinde getirmektedir. Zira oyun tabanlı bir testin arkasında çalışan algoritmanın, adayın bilişsel kapasitesini hangi ağırlıklı puanlama kriterlerine göre ölçtüğü veya adayın kişilik özelliklerine dair vardığı çıkarımların ne derece nesnel olduğu aday ve çoğu zaman işveren tarafından denetlenememektedir. Bu durum, adayın doğrudan işin niteliğiyle ilgisi olmayan

[8] Bkz. https://www.equalture.com/blog/psychometric-tests-in-recruitment-advantages-and-disadvantages/ [18.05.2026]

davranışsal bir refleks nedeniyle elenmesi riskini doğurarak, sözleşme öncesi süreçte tarafların eşitliği ve ayrımcılık yasağının sınırlarını yeniden tartışmaya açmaktadır[9].

8. İşe Yerleştirme

Yapay zekâ teorik olarak özerk seçim yapabilme kabiliyetine sahip olsa da günümüz kurumsal pratiklerinde esasen bir karar destek aracı olarak konumlandırılmaktadır. Özellikle iş sözleşmesinin kurulması gibi nihai hukuki aşamalarda insan müdahalesi ve muhakemesi hayati önemini korumaktadır. Bu süreçte büyük veri analitiği vasıtasıyla adayların işe uygunluğuna ilişkin kapsamlı profiller oluşturulmakta, *people analytics* uygulamaları ise aday seçimi ve performans tahmininde insan kaynakları departmanlarına stratejik karar desteği sunmaktadır. Tahmine dayalı puanlama modelleri işe yerleştirme sürecini hızlandırırken, veri temelli kararların güçlenmesini sağlamaktadır. Sektörde bu entegrasyonu sağlayan küresel araçlar incelendiğinde; *Canditech*'in video mülakatlarla birlikte yapay zekâ destekli beceri değerlendirmeleri sunduğu, *Workable*'ın en nitelikli adayları önceliklendirdiği, *Pinpoint*'in otomatik aday taraması gerçekleştirdiği, *Manatal*'ın yapay zekâ destekli aday bulma ve aday verisi zenginleştirme işlevini yürüttüğü görülmektedir. Ayrıca *Checkr* yapay zekâ destekli özgeçmiş taraması yaparken, *Humanly* yapay zekâ destekli aday etkileşimi kurmakta, *Boon* yapay zekâ odaklı aday yönlendirme ve referans sistemlerini işletmekte, *Recruit CRM* yapay zekâ destekli e-posta otomasyonu ve sıralı gönderim sağlamakta, *Syndio* yapay zekâ destekli ücret ve maaş önerileri geliştirmekte ve *Truffle* ise geniş aday havuzunu otonom yönetme imkânı tanımaktadır.

Son dönemde işe alım süreçlerinin yapay zekâ teknolojisi ile yürütülmesi gittikçe yaygın bir hale gelmiştir. Şirketler, yapılan yüzlerce başvuru arasından yapay zekâ aracılığı ile doğru adayın nesnel bir biçimde belirlenmesini hedeflemektedirler. Öyle ki, aday seçimi halihazırda insan kaynakları uzmanlarının temel görevleri arasında yer alırken, yapay zekâ entegrasyonu ile herhangi bir uzmanının insani doğası gereği yapabileceği hata payının sıfıra indirgenmesi amaçlanmaktadır. Nitekim yapay zekâ, insan emeğinin sınırlarını aşarak yüzlerce kişinin özgeçmişine aynı anda erişip bunlardan anlamlı bir çıktı oluşturabilmektedir. Türkiye'deki kurumsal pratikler incelendiğinde, *Turkcell*'in yeni mezun işe alım programının video mülakat aşamasında yaklaşık 2 bin adayın değerlendirilmesinde yapay zekâ algoritması kullandığı görülmektedir. Bu uygulamanın dikkat çekici yönü ise, program sonunda başarılı olan adaylara yönelik hukuki sonuç doğuran iş teklifinin de bizzat yapay zekâ aracılığıyla otonom olarak yapılmış olmasıdır. Benzer şekilde *Koç Holding* tarafından tasarlanan yapay zekâ yazılımı da öncelikli olarak Koç Topluluğu'nun kurumsal kariyer platformu olan *Koç Kariyerim*'de devreye alınmaktadır. Bu sistem, adayların öz geçmişlerini dijital ortamlardan otomatik oluşturarak veri tabanına işlemekte ve yeteneğe uygun bir pozisyon açıldığında aday henüz manuel bir başvuruda bulunmadan adayı değerlendirmeye alma imkânı sunmaktadır. Yılda 3 milyondan fazla başvuru alan Koç Topluluğu'nda, bu yazılım sayesinde hem insan gücü hem de zaman bakımından son derece önemli bir kurumsal kaynak tasarrufu sağlanması hedeflenmektedir.

Yapay zekânın sağladığı bu operasyonel ve mali avantajların yanı sıra, süreçlerin tamamen algoritmalara devredilmesi ciddi hak ihlalleri ve şeffaflık krizlerini de beraberinde getirmektedir. Bunun en güncel örneği, ABD'nin California eyaletinde; Microsoft, PayPal ve çok sayıda *Fortune 500* şirketi tarafından kullanılan yapay zekâ tabanlı işe alım platformu *Eightfold AI* hakkında açılan ilk davadır. Dava dilekçesinde, *Eightfold AI*'ın iş başvuru sahiplerinin bilgisi ve rızası olmadan onlar hakkında detaylı "değerlendirme raporları" hazırladığı ve bu raporlarda adaylara açıklanmayan "gizli uygunluk puanı" atadığı iddia edilmektedir. İşverenlerin adayları doğrudan elemek için kullandığı bu gizli puanlamalar; kişisel verilerin korunması hukuku bağlamında profil çıkarma sınırlarının aşılması ve veri sorumlusunun bilgilendirme/aydınlatma yükümlülüklerinin ihlali tartışmasını yargı gündemine taşımıştır[10].

Teknolojik dönüşüme yönelik bu eğilimler, endişeler ve sektörel refleksler, *Workable 2024 İşe Alım Sürecinde Yapay Zeka Anketi* verileriyle de ampirik olarak doğrulanmaktadır. Rapordaki öne çıkan önemli verilere göre, katılımcıların önemli bir çoğunluğu (%62,5) işe alım süreçlerinde aktif olarak yapay zekâ kullanmaktadır. Yapay zekânın işe alım sürecini oldukça hızlandırdığını düşünenlerin oranı %89,6 gibi yüksek bir seviyede iken; sistemin işe alım için harcanan süreyi %85,3, operasyonel maliyeti ise %77,9 oranında azalttığı tespit edilmiştir. İşe alımcıların çoğunluğu (%75,7) yapay zekânın kendi üretkenliklerini artırdığını söylese de, teknik zorluklar (%46,2) ve çalışan direnci (%40,5) gibi yapısal zorluklar varlığını korumaktadır. Nitekim çalışanların %68,1'i yapay zekânın işlerini gelecekte olumsuz etkilemesi konusunda ciddi endişeler taşımaktadır. İşe alımda yapay zekâ ile ilgili bildirilen en yaygın problemlerin başında %40 ile işe alma ön yargısı (algoritmik ayrımcılık) gelirken, bunu %37,2 ile gizlilik endişeleri ve %30,7 ile yapay zekâ sistemlerine uyum zorlukları takip etmektedir. Araştırma verilerine göre yapay zekâ işe alımda en çok özgeçmiş taraması (%58,9) ve aday eşleştirmesi (%43,1) süreçlerinde kullanılmaktadır. Ancak işe alımda "insan dokunuşu" önemini kaybetmemiştir; katılımcıların %56,8'i yapay zekayı yalnızca destekleyici bir araç olarak

[9] MIT Media Lab araştırmacılarından Joy Buolamwini, bazı yüz tanıma teknolojilerindeki önyargı riskini ortaya koymuştur. Araştırma kapsamında kullandığı yazılım, onun yüzünü ancak beyaz bir maske taktığında tanıyabilmiştir. Ayrıntılı için bkz. https://www.spiceworks.com/hr/future-work/guest-article/ai-bias-challenges-hr/ [18.05.2026]

[10] Bkz. https://www.reuters.com/sustainability/boards-policy-regulation/ai-company-eightfold-sued-helping-companies-secretly-score-job-seekers-2026-01-21/ [19.05.2026]

konumlandırmaktadır. Katılımcıların %15,3'ü işe alımda yapay zekâ yerine tamamen insan görüşünün öncelikli olduğunu belirtirken, %21,1'i ise insan görüşü ile yapay zekayı eşit ağırlıkta kullandığını ifade etmektedir. Sadece %6,7 gibi çok kısıtlı bir kesim işe alım kararlarını verirken daha çok yapay zekâya dayalı tavsiyelere güvenmektedir. Elde edilen tüm bu ampirik ve teorik veriler birlikte değerlendirildiğinde; yüksek işe alımcı görüşünün muhafaza edildiği ve yapay zekânın yalnızca destekleyici bir enstrüman olarak konumlandırıldığı hibrid modelin en uygun ve hukuken güvenli formül olduğu anlaşılmaktadır[11].

B. İşe Alımda Yapay Zekâ Kullanımının Avantajları ve Riskleri

İş gücü piyasasında yapay zekâ teknolojilerinin işe alım süreçlerine entegrasyonu, insan kaynakları yönetimi açısından hem operasyonel avantajları hem de çözülmesi gereken karmaşık hukuki ve etik riskleri eş zamanlı olarak barındırmaktadır. Dönüşümün sağladığı avantajlar boyutunda ilk olarak; operasyonel verimlilik, zaman ve maliyet tasarrufu öne çıkmaktadır. Yapay zekâ tabanlı sistemler, devasa başvuru havuzlarının çok kısa bir süre zarfında taranmasına imkân tanımakta; özellikle yüksek hacimli işe alım süreçlerinde insan kaynakları departmanlarının üzerindeki manuel tarama yükünü radikal bir biçimde azaltmaktadır. Bu durum, işe alım maliyetlerinin düşürülmesini ve toplam işe alım süresinin kısalmasını doğrudan beraberinde getirmektedir. Sürecin ikinci önemli avantajı ise standardizasyon ve tutarlılık zemininde somutlaşmaktadır. Ön değerlendirme kriterlerinin sistem tarafından objektif ve sistematik olarak uygulanması; adayın iş ilanıyla uyumu ve asgari nitelikleri taşıyıp taşımadığı gibi biçimsel konularda insan faktöründen kaynaklanabilecek hata payını asgariye indirmektedir. Bununla birlikte yapay zekâ, işe alımda ön yargıların azaltılması potansiyeline de sahiptir. Özellikle adayın fotoğrafını teknik olarak devre dışı bırakan ve isimleri tamamen anonimleştiren algoritmik sistemler sayesinde; ırk, cinsiyet ve yaş görünümü gibi unsurlara dayalı bilinç dışı toplumsal ön yargıların seçme süreçlerindeki ayrımcı etkileri büyük ölçüde engellenebilmektedir. Süreçlerin bu şekilde otonomlaşması, son olarak insan kaynaklarının kurumsal zaman ve kaynaklarını daha stratejik alanlara yönlendirmelerine fırsat tanımaktadır. İnsan kaynakları uzmanları, zaman alan mekanik ve idari işlerle boğuşmak yerine; aday deneyiminin iyileştirilmesi, stratejik iş gücü planlaması ve iç yetenek havuzu yönetimi gibi kurumsal katma değeri yüksek, vizyoner alanlara odaklanma imkânı bulmaktadır.

Madalyonun diğer yüzünde yer alan hukuki ve etik riskler ise, algoritmik sistemlerin iş hukuku doktrininde ve temel haklar boyutunda yarattığı gri alanları gözler önüne sermektedir. Bu risklerin başında, geçmiş veri setlerindeki kalıplaşmış "örüntülerin" yapay zekâ tarafından öğrenilmesi sonucu ortaya çıkan algoritmik ayrımcılık gelmektedir. Sistemin eğitildiği tarihsel veri setlerindeki gizli yanlılıklar ve ön yargılı veri modelleri, algoritmalar tarafından sistematik birer eleme kriterine dönüştürülerek ayrımcılığın dijital ortamda yeniden üretilmesine yol açabilmektedir. Bu yapısal soruna, karar alma mekanizmalarındaki şeffaflık eksikliği (kara kutu problemi) eşlik etmektedir. Yapay zekâ tarafından elenen bir aday; süreç boyunca kendi hakkında hangi kişisel verilerin işlendiğini, değerlendirme puanının tam olarak nasıl belirlendiğini ve en nihayetinde hangi somut kritere dayanılarak elendiğini bilememekte, bu durum da kararların denetlenebilirliğini ve hak arama özgürlüğünü zedelemektedir. Sürecin barındırdığı bir diğer teknik ve hukuki handikap ise veri güvenliği ve siber riskler alanında ortaya çıkmaktadır. Adaylara ait hassas nitelikteki kişisel bilgilerin, profilleme amacıyla devasa veri setlerinde depolanması yapısal bir güvenlik açığı yaratmaktadır; zira yapay zekâ sistemleri doğası gereği yüksek hacimli veri barındırdığından, siber saldırıların ve veri sızıntılarının yaratacağı hak ihlali riski de geometrik olarak artmaktadır. Bunlara ek olarak, özellikle adayların sosyal medya mecralarındaki dijital ayak izlerinin ve çevrimiçi verilerinin rıza dışı kullanımı, Kişisel Verilerin Korunması Kanunu (KVKK) ile anayasal bir hak olan özel hayatın gizliliği hakkı bakımından ciddi ihlal riskleri doğurmaktadır.

Sürecin felsefi ve insan odaklı boyutunda ise insan onuru ve öznellik kavramları zedelenmektedir. Yapay zekâ mimarileri; insan davranışlarının altında yatan sosyal bağlamı, kültürel nüansları ve yüz yüze etkileşimdeki kişisel iletişimin derinliğini rasyonel olarak değerlendirebilecek kapasiteye sahip değildir. Dolayısıyla, aşırı derecede mekanikleştirilen işe alım süreçleri, bireyin bir nesneye indirgenmesi riski nedeniyle temel insan onuru ilkeleriyle açıkça çelişebilmektedir. Sistemlerin empati, adalet kavramı ve etik değerlendirme yapabilme kabiliyeti bakımından yapısal sınırlılıkları bulunduğundan, yapay zekânın mülakat ve potansiyel değerlendirmelerinde insan muhakemesinin yerini tam anlamıyla alması imkânsız görünmektedir. Tüm bu işlevsel faydalar ile çok boyutlu risk faktörleri birlikte tartıldığında; işe alım prosedürlerinin geleceği açısından en optimum, etik ve hukuken savunulabilir yaklaşımın, yapay zekânın hız ve tarama gücünden yararlanırken nihai kararda insan denetimini ve vicdanını muhafaza eden "insan ve yapay zekâ hibrit karar modeli" olduğu anlaşılmaktadır.

[11] Bkz. https://resources.workable.com/stories-and-insights/five-challenges-facing-managers-in-mid-2024 [19.05.2026]

C. İşe Alımda Yapay Zekâ Kullanımından Doğabilecek Sorunlar

1. Algoritmik Ayrımcılık Riski

Yapay zekâ destekli işe alım süreçlerinin en kritik yapısal handikaplarından birini oluşturan algoritmik ayrımcılık riski; bir algoritmanın ya da yapay zekâ sisteminin, belirli kişilere ya da toplumsal gruplara karşı sistematik olarak ve haksız bir biçimde olumsuz sonuçlar üretmesini ifade etmektedir[7], [8]. Bu sistematik dışlanma probleminin temel kaynağını, algoritmanın eğitildiği tarihsel veri setlerindeki ön yargılar ya da sistemsel tasarım hataları oluşturmaktadır[6],[9]. Bir başka ifadeyle, yapay zekâ mimarileri geçmiş dönemin insan kaynakları kararlarındaki bilinçli ya da bilinç dışı ayrımcı örüntüleri öğrenerek geleceğe yönelik kararlarında bunları rasyonel kurallarmış gibi tatbik etmektedir[10], [11], [12]. İşe alım sürecinde yapay zekâ teknolojilerinin neden olduğu bu algoritmik ayrımcılık, hem adaylar hem de işverenler açısından son derece olumsuz sonuçlar yaratabilmektedir; adaylar bakımından anayasal çalışma hakkı ve fırsat eşitliği zedelenirken, işverenler açısından ise liyakatli yeteneklerin kaybına ve en önemlisi, eşit işlem yapma borcunun ihlali nedeniyle ciddi hukuki sorumlulukların doğmasına sebebiyet vermektedir[13].

Algoritmik ayrımcılık riskinin kurumsal süreçlerdeki yansımaları, iş hukukunun ve temel insan haklarının koruma altına aldığı belirli korumalı kategoriler üzerinde somutlaşmaktadır[14]. Bu kategorilerin başında gelen cinsiyet temelli ayrımcılık, yapay zekâ algoritmalarının geçmiş dönem iş deneyimi, sektörel istihdam yoğunluğu ve eğitim verileri gibi özelliklere dayanarak adayları değerlendirirken, toplumsal cinsiyet rolleri gibi ön yargılı faktörleri de veri modeline dahil etmesiyle ortaya çıkmaktadır. Bu durumda sistem, tarihsel olarak erkek egemen olan belirli pozisyonlar veya üst düzey yöneticilik roller için erkek adaylara matematiksel olarak daha fazla ağırlık verebilmekte; kadın adayları ise özgeçmişlerindeki dil kalıpları veya kariyer boşlukları üzerinden ön yargılı bir şekilde elemine edebilmektedir.

Benzer bir yapısal sapma, etnik köken temelli ayrımcılık süreçlerinde de gözlenmektedir. Yapay zekâ tabanlı sistemler, adayların doğrudan beyan etmedikleri durumlarda dahi, isim ve soyisim kombinasyonlarından, mezun olunan okullardan veya ikamet edilen coğrafi bölgelerden yola çıkarak adayların etnik kökenlerini tahmine dayalı yöntemlerle belirleyebilmektedir. Bu algoritmik çıkarım neticesinde, belirli etnik gruplara veya azınlıklara mensup adaylar sistem tarafından sistematik olarak daha düşük uygunluk puanlarına maruz bırakılabilmekte ve böylelikle işe alım sürecinde doğrudan ya da dolaylı ayrımcılığa uğrayabilmektedirler.

Bunun yanı sıra, iş gücü piyasasının sürdürülebilirliği önündeki en büyük engellerden biri olan yaş temelli ayrımcılık da dijitalleşmeyle birlikte yeni bir boyut kazanmıştır. Yapay zekâ; adayların doğum tarihleri, mezuniyet yılları veya geçmiş iş deneyimlerinin kronolojik süresi gibi nesnel görünen verilere dayanarak adayların yaş gruplarını kolaylıkla tahmin edebilmektedir. Bu algoritmik modelleme sonucunda, belirli bir yaşın üzerindeki kıdemli adaylar veya kariyerinin başındaki genç profesyoneller, ön yargılı maliyet ve performans algıları nedeniyle belirli pozisyonlar için peşinen daha az uygun görülebilmekte ve işe alım sürecinde fırsat eşitliğinden mahrum bırakılabilmektedirler.

Son olarak, kırılgan grupların istihdama katılımı önündeki en hassas bariyerlerden birini engellilik temelli ayrımcılık riski oluşturmaktadır. Yapay zekâ mekanizmaları; adayların sağlık geçmişine ilişkin dolaylı verileri, sigorta kayıtlarındaki boşlukları veya geçmiş iş gücü geçmişindeki dönemsel kesintileri analiz ederek adayların engellilik veya kronik rahatsızlık durumlarını otonom olarak tahmin edebilmektedir. Bu durumda, engelli adayların potansiyel performansları veya adaptasyon süreçleri hakkında ön yargılı varsayımlarda bulunan algoritmalar, bu adayları belirli açık pozisyonlar için peşinen dezavantajlı olarak kodlayabilmekte ve adayların mülakat aşamasına dahi geçemeden sistem dışına itilmesine yol açabilmektedir.

Teorik düzeyde tartışılan bu algoritmik ve sistematik ayrımcılık riskleri, küresel teknoloji devlerinin uygulamaya koyduğu yapay zekâ sistemlerinde somut hukuki uyuşmazlıklara ve operasyonel başarısızlıklara dönüşmüştür. Bu durumun dünya çapındaki en bilinen örneklerinden biri, e-ticaret firması Amazon'un işe alım süreçlerini otomatikleştirmek amacıyla tasarladığı yapay zekâ mekanizmasının cinsiyetçi olduğu gerekçesiyle kapatılmasıdır[15], [16]. ABD merkezli firma, geliştirdiği makine öğrenimi modelinin açıkça toplumsal cinsiyet temelinde ayrımcı sonuçlar doğurduğunu saptamış ve dört yıl boyunca işe alımlarda aktif olarak kullandığı bu otonom sistemden 2018 yılında tamamen vazgeçmek zorunda kalmıştır. Amazon vakasında yapay zekânın bu denli büyük bir algoritmik sapma yaşamasının nedeni, sistemin veri toplama ve model geliştirme aşamasında şirkete geçmişte yapılmış olan 10 yıllık özgeçmiş havuzunu referans alarak eğitilmiş olmasıdır. Söz konusu tarihsel süreçte teknoloji ve yazılım sektöründe erkek iş gücünün ezici bir üstünlüğe ve egemenliğe sahip olması nedeniyle, yapay zekâ bu örüntüyü nesnel bir başarı kriteri olarak kodlamış; neticede "erkek personelin kadın personele kıyasla daha tercih edilebilir" olduğu yönünde hatalı ve ön yargılı bir yargıya vararak kadın adayların özgeçmişlerini sistematik olarak elemiştir.

Benzer şekilde, hedefli reklamcılık ve ilan yayımı süreçlerinde algoritmaların kullanımı, Meta (Facebook) şirketine yönelik ciddi hukuki şikayetleri ve yeni davaları beraberinde getirmiştir[17],[18]. Meta, şirketlerin platform üzerinde iş ilanlarıyla potansiyel aday kullanıcıları hedeflemek için kullandığı dağıtım algoritmasının yapısal olarak ayrımcı olduğu iddiasıyla çok sayıda yasal şikayetle karşı karşıya kalmıştır. Yapılan incelemelerde, Facebook'un reklam dağıtım algoritmasının iş kollarına yönelik toplumsal ön yargıları yapay zekâ eliyle derinleştirdiği tespit edilmiştir. Örneğin; mekanik

tamirat veya teknik uzmanlık gerektiren tamirci pozisyonlarına ilişkin iş ilanları ezici bir çoğunlukla erkek kullanıcıların dijital ekranlarına düşürülürken, okul öncesi öğretmenliği veya çocuk bakımı gibi bakım emeğine dayalı mesleklerin reklamları çoğunlukla kadın kullanıcılara gösterilmiştir. Bu algoritmik yönlendirmenin daha da vahim boyutu ise ırksal ve ekonomik ayrımcılığı tetiklemesidir; nitekim sistem, beyaz erkek kullanıcıların karşısına geliri ve kariyer statüsü çok daha yüksek iş ilanlarını çıkarırken, kadınlar veya beyaz ırktan olmayan (azınlık/göçmen) kullanıcılar içinse sistematik olarak daha düşük ücretli ve güvencesiz iş ilanlarını görünür kılmıştır.

Algoritmik ayrımcılığın iş hukuku, ispat hukuku ve fikri mülkiyet hukuku kavşağında yarattığı en karmaşık krizlerden biri de kurumsal yazılım devi Workday aleyhine açılan davada somutlaşmıştır[19], [20]. Amerika Birleşik Devletleri'nde görülen davada Workday sistemi; siyahi, belirli bir yaşın üzerindeki yaşlı ve engelli adayları daha ilk tarama aşamasında otonom olarak ve otomatik bir biçimde elemekle suçlanmıştır. Mağdur adaylar, işe alım platformunun ve kullanılan yapay zekânın taraflı, kapsayıcı olmayan ve ön yargılı tarihsel veri setleriyle eğitildiğini ileri sürmüşlerdir. Workday davası, akademik literatürde "ispat yükü" ile "ticari sır ve şeffaflık dengesi" arasındaki asimetriyi tartışmaya açması bakımından kritiktir. Zira mevcut hukuk sistemlerinde ayrımcılığa uğradığını iddia eden aday, bu iddiasını ispat etmekle yükümlüyken; elenmesine neden olan algoritmanın kaynak kodları, ağırlık parametreleri ve veri setleri yazılım şirketleri tarafından "ticari sır" ve fikri mülkiyet koruması arkasına gizlenmektedir. Bu durum, adaylar ve işçiler bakımından tam bir korumasızlık ve şeffaflık krizine yol açarak, dijitalleşen işe alım süreçlerinde adil yargılanma ve eşit işlem ilkelerinin denetimini imkânsız hale getiren yapısal bir bariyer oluşturmaktadır.

2. Veri Gizliliği İhlali Riski

Yapay zekâ destekli işe alım sistemleri, doğaları gereği çok boyutlu kişisel veri işleme faaliyetlerinin merkezinde yer almaktadır[11]. Bu alanda kullanılan algoritmik uygulamaların büyük bir kısmı, adayların dijital ayak izlerini takip ederek otomatik analiz yapma ve profil çıkarma niteliği taşımaktadır. Sistemin adayların dil kullanımlarından, mülakat esnasındaki jest ve mimiklerinden ya da sosyal medya etkileşimlerinden yola çıkarak gerçekleştirdiği davranışsal ve kişilik temelli çıkarımlar, veri koruma hukuku bakımından yönetilmesi güç ve son derece kritik olan "özel nitelikli kişisel veri" işleme riskini doğurmaktadır[21]. Pozitif hukuk metinlerinde açıkça düzenlendiği üzere; kişilerin ırkı, etnik kökeni, siyasi düşüncesi, felsefi inancı, dini, mezhebi veya diğer inançları, kılık ve kıyafeti, dernek, vakıf ya da sendika üyeliği, sağlığı, cinsel hayatı, ceza mahkûmiyeti ve güvenlik tedbirleriyle ilgili verileri ile biyometrik ve genetik verileri özel nitelikli kişisel veri kapsamında kabul edilmektedir[22]. Bu hassas verilerin hukuka aykırı şekilde işlenmesi, ifşa edilmesi veya algoritmik bir modelin girdisi haline getirilmesi durumunda, veri sahibi olan adayın telafisi imkânsız zararlara uğraması ya da iş gücü piyasasında doğrudan ayrımcı muameleye maruz kalması kaçınılmazdır[23], [24].

Veri koruma hukuku bağlamında asıl hukuki karmaşayı ve gri alanı yaratan unsur ise, yapay zekâ sistemlerinin çıkarımsal yetenekleridir. Gelişmiş makine öğrenimi algoritmaları, nihai bir çıktı elde etme sürecinde, başlangıçta kendisine sağlanan sıradan ve genel nitelikli kişisel verilerden (örneğin adayın hobileri, üye olduğu mesleki topluluklar, paylaştığı metinlerin dil kalıpları veya seyahat sıklığı) yola çıkarak, veri sahibinin özel nitelikli kişisel verilerine ilişkin de dolaylı tahminler ve çıktılar üretebilmektedir. Yapay zekânın, adayın doğrudan beyan etmediği sendikal eğilimlerini, dini inançlarını, sağlık geçmişini veya cinsel yönelimini korelasyonel analizlerle "tahmin etmesi" ve bu çıkarımları işe alım puanlamasında gizli birer parametre olarak kullanması, açık bir veri gizliliği ihlalidir. Dolayısıyla, işe alım süreçlerinde şeffaf ve denetlenebilir olmayan bir yapay zekâ modelinin kullanılması; adayların rıza sınırlarını aşan, veri işlemedeki dürüstlük ve amaçla sınırlılık ilkelerini zedeleyen ve veri koruma mevzuatının emredici hükümleriyle doğrudan çatışan yapısal bir risk barındırmaktadır.

İşe alım süreçlerinde yapay zekâ teknolojilerinin entegrasyonu, doğası gereği son derece yüksek düzeyde bir kişisel veri işleme faaliyetini bünyesinde barındırmaktadır. Bu süreçte algoritmaların gerçekleştirdiği "profil çıkarma", "otomatik karar verme" ve mülakatlar esnasında başvurulan "biyometrik veri işleme" gibi spesifik operasyonlar bakımından 6698 sayılı Kişisel Verilerin Korunması Kanunu (KVKK)[12] hükümleri doğrudan uygulama alanı bulmaktadır. Kanun'un 5. ve 6. maddelerinde düzenlenen veri işlemenin hukuka uygunluk şartları çerçevesinde; adayların ırkı, etnik kökeni veya sağlığı gibi hassas bilgileri içeren özel nitelikli kişisel verileri, kanunlarda açıkça öngörülen istisnai haller dışında, ancak adayın üzerinde hiçbir baskı unsuru hissetmeden özgür iradesiyle açıklayacağı "açık rızası" ile işlenebilmektedir. Bu bağlamda, yapay zekâ sistemlerinin asenkron video mülakatlarda sıklıkla başvurduğu duygu analizi, yüz tanıma, mikro mimik takibi ve ses tonu frekansı ölçümü gibi ileri düzey teknik yöntemler, veri koruma hukuku tahtında doğrudan "biyometrik veri" niteliği taşımaktadır. Dolayısıyla iş ilişkisinin kurulmasından önceki sözleşme öncesi süreçte, adayların bu tür biyometrik verilerinin geçerli bir açık rıza mekanizması işletilmeden otonom algoritmalar tarafından toplanması ve işlenmesi açık bir hukuka aykırılık teşkil etmektedir.

Öte yandan, KVKK'nın 4. maddesinde kişisel veri işlemenin temel ilkeleri olarak düzenlenen "veri minimizasyonu" (ölçülülük) ve "amaçla sınırlılık" prensipleri de yapay zekâ

[12] Kişisel Verilerin Korunması Kanunu, Kanun No. 6698, Resmî Gazete, no. 29677, 7 Nisan 2016.

mimarileriyle yapısal bir çatışma içerisindedir. Makine öğrenimi tabanlı yapay zekâ sistemleri, doğaları gereği daha doğru tahmin modelleri geliştirebilmek için büyük veri yığınlarına ihtiyaç duymakta ve çoğu zaman açık pozisyonun niteliğiyle ilgisi bulunmayan, gereğinden fazla veriyi adaylardan talep etmektedir. Adayın işe uygunluğunu ölçmek için doğrudan gerekli olmayan bu aşırı veri toplama eğilimi, Kanun'un dürüstlük kuralına uygun olma, işlendikleri amaçla bağlantılı, sınırlı ve ölçülü olma ilkelerine doğrudan aykırılık teşkil etmektedir. Veri sahibinin haklarının koruma altına alınması noktasında asıl güvenceyi ise KVKK'nın 11. maddesi oluşturmaktadır. İlgili madde uyarınca adayların, "*işlenen verilerin münhasıran otomatik sistemler vasıtasıyla analiz edilmesi suretiyle kendileri aleyhine bir sonucun ortaya çıkmasına itiraz etme"* hakkı söz konusudur. Yapay zekânın özgeçmiş taraması veya mülakat puanlaması neticesinde adayı otomatik olarak elemesi, aday aleyhine doğmuş net bir hukuki sonuçtur. Bu kapsamda aday, kanundan doğan bu hakkını ileri sürerek elenme kararının tamamen otonom sistemlerin tekelinden çıkarılmasını ve kararın tarafsız bir uzman tarafından yeniden değerlendirilmesini, yani süreç üzerinde mutlak bir "insan denetimi" (*human-in-the-loop*) gerçekleştirilmesini işverenden talep etme yetkisine sahiptir.

Avrupa Birliği Genel Veri Koruma Tüzüğü (GDPR), küresel ölçekte işe alım süreçlerinde kişisel veri işlemenin hukuka uygunluğunu ve otomatik karar vermenin yasal sınırlarını belirleyen en temel ve emredici normatif metin niteliğindedir[13]. Tüzük'ün 6. maddesi veri işlemenin hukuka uygunluk sebeplerini (sözleşmenin kurulması, veri sorumlusunun meşru menfaati, rıza vb.) düzenlerken; 9. maddesi ise adayların hassas verilerinin işlenmesini çok daha sıkı koruma şartlarına bağlamaktadır. Bununla birlikte, veri sorumlusunun aydınlatma yükümlülüğünü ve ilgili kişinin bilgi edinme hakkını düzenleyen 13, 14 ve 15. maddeler ile veri sahibinin profil çıkarma faaliyetlerine karşı itiraz hakkını saklı tutan 21. madde, yapay zekâ destekli işe alım süreçlerinin şeffaflık zemininde yürütülmesini zorunlu kılmaktadır. Bu kapsamda 13. madde veri sahibinden kişisel verilerin doğrudan temin edildiği hallerde sağlanması gereken bilgileri, 14. madde kişisel verilerin veri sahibinden alınmadığı (üçüncü taraf veya dijital ağlardan tarandığı) hallerde sağlanacak bilgileri, 15. madde ise veri sahibinin kendi verilerine yönelik genel erişim hakkını hüküm altına almaktadır. Bu düzenlemelerin ortak amacı uyarınca; veri sahibinin, profil çıkarma da dahil olmak üzere otomatik karar vermenin varlığı hakkında bilgilendirilme hakkı açıkça mevcuttur. Aday, en azından bu hallerde, sistemde yürütülen algoritmik mantığa ilişkin anlamlı bilgilerin yanı sıra söz konusu veri işleme faaliyetinin kendisi açısından taşıdığı önemi ve öngörülen sonuçları hakkında bilgi edinme ve erişim hakkına sahiptir.

GDPR'ın işe alım algoritmaları karşısında önemli bir koruma hükmünü ise Tüzük'ün 22. maddesi oluşturmaktadır. İlgili madde, kural olarak, ilgili kişiyi hukuken veya benzer şekilde önemli ölçüde etkileyen ve "yalnızca otomatik işleme" esasına dayanan karar verme mekanizmalarını açıkça yasaklamaktadır. Madde 22'nin 1. fıkrasında yer alan *"Veri sahibi, yalnızca otomatik işleme, dahil olmak üzere profil oluşturmaya dayalı, kendisini önemli ölçüde etkileyen veya benzer hukuki sonuçlar doğuran bir karara tabi tutulmama hakkına sahiptir"* emredici hükmü, otonom sistemlerin bireyler üzerindeki mutlak egemenliğini sınırlandırmayı amaçlamaktadır. İşe alım prosedürleri, bireyin doğrudan ekonomik geleceğini, kariyerini ve sosyal hayatını şekillendiren bir nitelik taşıdığı için, 22. madde anlamındaki "benzer şekilde önemli ölçüde etkileme" kriterini şüpheye yer bırakmayacak biçimde karşılamaktadır. Bu doğrultuda, bir adayın insan kaynakları süreçlerinde sadece bir yapay zekâ algoritmasının verdiği kararla doğrudan reddedilmesi veya kabul edilmesi, Avrupa veri koruma hukuku tahtında kural olarak yasaklanmıştır.

Söz konusu yasal yasağın, sözleşmenin kurulması için zorunluluk arz etmesi veya adayın açık rızasının bulunması gibi istisnai hallerde esnetilmesi durumunda dahi, adaya çok güçlü güvenceler tanınmıştır. Bu istisnai durumlarda bile ilgili kişiye, yapay zekânın verdiği karara esastan itiraz etme, kendi görüşünü sunma ve en önemlisi sürecin arkasında mutlak surette bir "insan incelemesi" talep etme hakkı tanınmalıdır. Bu durum, bireylerin kendileri üzerinde önemli hukuki ve ekonomik etkileri olabilecek kararların alınmasında insan müdahalesi olmasını talep etme hakkına sahip olmaları anlamına gelir. Dolayısıyla aday, işe alım sürecinde kullanılan karmaşık algoritmaların temel işleyiş mantığı, bu yapay zekâ kararının doğuracağı hukuki sonuçlar ve sürecin bizzat kendisini nasıl etkilediği konusunda açık, şeffaf ve "anlaşılır bilgi" alma hakkına sahiptir. Bu yapı, işverenin yönetim hakkı ile adayın kişilik hakları ve veri güvenliği arasında kurulması gereken adil dengenin yasal çerçevesini çizmektedir.

İşe alımda yapay zekâ ve büyük veri analitiğinin doğurduğu hukuki ve etik risklerin bertaraf edilmesi, adayların temel haklarının korunması bakımından elzemdir. Bu kapsamda geliştirilmesi gereken ilk strateji mahremiyetin korunması ilkesidir. İşverenler, "tasarımdan itibaren mahremiyet" (*privacy by design*) yaklaşımını benimseyerek, toplanan verileri sadece açık pozisyonun niteliğiyle sınırlı tutmalı (veri minimizasyonu) ve aday havuzundaki verilerin maskelenmesi gibi teknik tedbirleri en baştan itibaren sisteme entegre etmelidir. İkinci olarak, algoritmaların birer "kara kutu" gibi işlemesini önlemek adına şeffaflığın sağlanması zorunludur. Adayların, elenme gerekçelerini, puanlama modellerini ve sistemin arkasında çalışan mantıksal örüntüyü anlaşılır bir dille öğrenebilecekleri şeffaf aydınlatma mekanizmaları kurulmalıdır.

[13] Avrupa Birliği Genel Veri Koruma Tüzüğü (General Data Protection Regulation–GDPR), Regulation (EU) 2016/679, Official Journal of the European Union, Apr. 27, 2016.

Sürecin anayasal boyutu için algoritmik modellerin ayrımcılık yasağına uyulması kriteri ekseninde düzenli denetlenmesi gerekmektedir. Yazılım algoritmaları, cinsiyet, etnik köken, yaş veya engellilik temelinde dolaylı ayrımcılığa yol açabilecek ön yargılı veri setlerinden arındırılmalı; periyodik "algoritmik etki değerlendirmeleri" yapılmalıdır. Teknik boyutta ise, adaylara ait yüksek hacimli biyometrik ve kişisel verilerin depolandığı veri tabanlarının siber saldırılara karşı korunması amacıyla en üst düzey siber güvenlik önlemlerinin alınması gerekir. Verilerin şifrelenmesi ve erişim yetkilerinin sınırlandırılması bu ihlal risklerini asgariye indirecektir.

Son olarak, tüm bu teknik sürecin üzerinde mutlak bir insan gözetimi yapılması şarttır. Yapay zekâ, empati, hakkaniyet ve kültürel bağlam gibi insani unsurları değerlendirme yeteneğinden yoksundur. Bu nedenle algoritmalar hiçbir zaman tek başına nihai bir eleme ya da kabul kararı vermemeli, yalnızca birer karar destek aracı olarak konumlandırılmalıdır. Otomatik sistemlerin ürettiği çıktılar, her halükârda insan kaynakları uzmanlarının veya yöneticilerin nihai onayından geçmelidir. Ancak bu hibrid model ve bütüncül önlemler sayesinde yapay zekânın verimliliğinden yararlanılırken adayların hakları da güvenceye kavuşturulabilir.

3. Tam Gözetim Yasağı ve Kara Kutu Sorunu

Yapay zekâ teknolojilerinin işe alım süreçlerindeki ağırlığının artması, hukuk doktrininde **"**tam gözetim yasağı" (*prohibition of total surveillance*) ve "kara kutu" (*black box*) sorunu ekseninde yeni bir tartışma alanı açmıştır. İleri düzey yapay zekâ uygulamaları, adayların işe uygunluğunu ölçme gayesinin ötesine geçerek bireyin tüm dijital varlığını kuşatan bir "tam gözetim" mekanizmasına dönüşmesi yönüyle sert eleştirilere konu olmaktadır. Adayların özel e-posta kutularının ve gizli iletişim kanallarının otonom yazılımlarla analiz edilmesi, sosyal medya etkileşimlerinden ve dijital ayak izlerinden yola çıkılarak derin davranışsal profillerinin çıkarılması, mülakatlar esnasında uygulanan duygu analizi ve ruh hâli tahminleme yöntemleri bu gözetimin boyutlarını ortaya koymaktadır[25]. Sözleşme öncesi süreçte (işe alım aşamasında) işverenin yönetim yetkisini aşan bu orantısız uygulamalar, adayların anayasal güvence altındaki özel hayatının gizliliği hakkı ve bireylerin dijital ortamlarda kendilerini özgürce ifade etmelerini engelleyen bir otosansür mekanizmasını tetiklemesi yönüyle ifade özgürlüğü üzerinde ağır bir baskı yaratmaktadır.

Bu hukuksal risklerin teknik boyuttaki en büyük çıkmazı ise, özellikle derin öğrenme (*deep learning*) ve yapay sinir ağları temelli sistemlerin yapısal olarak sahip olduğu "kara kutu" niteliğidir[26]. Bu gelişmiş algoritmik modeller, milyarlarca parametre arasındaki doğrusal olmayan karmaşık ilişkileri ve korelasyonları işleyerek kararlar ürettiği için, sistemin belirli bir adayı neden elediği veya hangi kriteri başarı unsuru saydığı nesnel olarak geriye dönük açıklanamamaktadır. Kararların gerekçelendirilememesi ve algoritmik mantığın şeffaf bir biçimde ortaya konulamaması, işe alım sürecinde dürüstlük kuralını ve eşit işlem yapma borcunu zedelediği gibi; adayların elenme kararlarına karşı hukuki yollara başvurmasını ve uğradıkları hak ihlallerini ispat etmelerini imkânsız hale getiren yapısal bir denetimsizlik alanı doğurmaktadır.

IV. AVRUPA BİRLİĞİ YAPAY ZEKÂ TÜZÜĞÜ

Yapay zekâ teknolojilerinin küresel ölçekteki hukuki çerçevesini radikal bir biçimde şekillendiren en güncel ve somut yasal gelişme, Avrupa Birliği bünyesinde hayata geçirilen Yapay Zekâ Tüzüğü'dür (*AI Act*) [14] . Avrupa Komisyonu tarafından ilk olarak 21 Nisan 2021 tarihinde sunulan ve yapay zekâ sistemlerinin piyasaya arzı, hizmete sunulması ile temel hakları tehdit eden bazı yıkıcı uygulamaların yasaklanmasına dair emredici kuralları belirleyen "Yapay Zeka Hakkında Uyumlaştırılmış Kurallar Getiren ve Bazı Birlik Yasama Tasarruflarını Değiştiren (AB) 2024/1689 sayılı Tüzük", 12 Temmuz 2024 tarihinde AB Resmi Gazetesi'nde yayımlanmıştır. Söz konusu tarihi tüzük, 1 Ağustos 2024 tarihinde resmen yürürlüğe girmiş olup, getirdiği yükümlülüklerin karmaşıklığı ve sektörel adaptasyon süreçleri dikkate alınarak, farklı risk kategorilerine ilişkin tüm hükümlerinin 2025 yılının Şubat ayından başlayarak 2030 yılı sonuna dek kademeli ve aşamalı bir takvim dahilinde uygulanması kararlaştırılmıştır. Avrupa Birliği'nin bu kapsamlı düzenlemesi, bir yandan dijital iç pazarın işleyişinin geliştirilmesini ve yenilikçiliğin desteklenmesini hedeflerken; diğer yandan insan odaklı, etik ve güvenilir yapay zekâ teknolojilerinin benimsenmesini teşvik etmeyi amaçlamaktadır. Birlik bu yasal hamleyle, yapay zekâ sistemlerinin olası zararlı ve öngörülemez etkilerine karşı toplumsal sağlığın, siber güvenliğin ve çevrenin korunmasının yanı sıra, modern demokrasinin ve hukukun üstünlüğünün en temel direği olan kişilik hakları ile temel hak ve özgürlüklerin korunmasını güvence altına almayı amaçlayan evrensel bir koruma kalkanı inşa etmektedir.

Avrupa Birliği Yapay Zekâ Tüzüğü, getirdiği emredici yükümlülükleri ve yasal sınırları belirlerken, yapay zekâ sistemlerinin toplum, birey ve temel haklar üzerinde yaratabileceği olası zararları esas alan "risk temelli bir yaklaşım" (*risk-based approach*) benimsemiştir. Bu doğrultuda tüzük kapsamında, teknolojilerin tasarım ve kullanım amaçlarına göre dört temel risk kategorisi belirlenmiştir:

Kabul Edilemez Risk (*Unacceptable Risk*): İnsan onuru, demokrasi ve temel haklar üzerinde açık bir tehdit oluşturan yapay zekâ uygulamaları bu kategoride yer almaktadır. Bireylerin sosyal davranışlarına göre puanlanmasını ifade eden kamusal "sosyal skorlama" (*social scoring*) sistemleri, kişilerin rızası olmaksızın biyometrik verilerle kitlesel olarak izlenmesi veya bireylerin özgür iradesini manipüle eden bilinçaltı

[14] Artificial Intelligence Act (AI Act), Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence, Official Journal of the European Union, Jul. 12, 2024.

tekniklerin kullanımı bu kapsama girmektedir. Tüzük uyarınca, kabul edilemez risk kategorisindeki yapay zekâ sistemlerinin Avrupa Birliği pazarında arzı, hizmete sunulması ve kullanımı kesin olarak yasaklanmıştır.

Yüksek Risk (*High Risk*): Bireylerin sağlığı, güvenliği veya temel hakları üzerinde ciddi olumsuz etkiler yaratma potansiyeli taşıyan sistemler bu grupta sınıflandırılmaktadır. Kritik altyapı yönetimi, eğitim ve mesleki eğitim süreçleri, kolluk kuvvetleri uygulamaları, adalet mekanizmaları ve özellikle işe alım ile insan kaynakları yönetimi süreçlerinde kullanılan tüm otonom seçme, tarama ve değerlendirme algoritmaları doğrudan yüksek riskli yapay zekâ sistemi olarak kabul edilmektedir. Bu kategorideki sistemlerin piyasaya sürülmeden önce ve kullanım süresince; risk yönetim sistemlerinin kurulması, yüksek kaliteli veri setlerinin kullanımı (ayrımcılığı önlemek adına), detaylı teknik dokümantasyon, şeffaflık, siber güvenlik standartları ve mutlak bir insan gözetimi sağlama gibi son derece katı ve emredici uygunluk yükümlülüklerini yerine getirmesi zorunludur.

Sınırlı Risk (*Limited Risk*): Bireyler üzerinde doğrudan tehlike arz etmeyen ancak manipülasyon veya yanıltma potansiyeli barındıran yapay zekâ uygulamalarını kapsamaktadır. Sınırlı risk grubundaki sistemler için tüzük, ağırlıklı olarak şeffaflık yükümlülükleri getirmektedir. Bu doğrultuda, kullanıcıların bir insanla değil, bir yapay zekâ sistemiyle etkileşimde bulundukları konusunda açık ve anlaşılır bir biçimde bilgilendirilmeleri yasal bir zorunluluktur.

Minimum Risk (*Minimal Risk*): Toplum veya bireysel haklar üzerinde hiçbir güvenlik riski ya da hak ihlali tehdidi barındırmayan sistemlerdir. Yapay zekâ destekli video oyunları, spam filtreleme yazılımları veya basit veri işleme algoritmaları bu kategoride yer alır. Tüzük, piyasadaki yapay zekâ uygulamalarının ezici bir çoğunluğunu oluşturan bu grup için herhangi bir ek yasal yükümlülük veya kısıtlama getirmemekte; bu sistemlerin mevcut genel hukuk kuralları çerçevesinde serbestçe dolaşımına ve gelişimine izin vermektedir.

Avrupa Birliği Yapay Zekâ Tüzüğü, getirdiği emredici yükümlülükleri ve yasal sınırları belirlerken, sistemlerin toplum ve temel haklar üzerinde yaratabileceği olası zararları esas alan risk temelli bir yaklaşım benimsemiştir. Bu doğrultuda Tüzük m. 5 uyarınca düzenlenen "kabul edilemez risk" kategorisindeki uygulamalar sınırlı sayı ilkesine (*numerus clausus*) tabi tutularak açıkça yasaklanmıştır. Yasaklanan bu yapay zekâ uygulamaları arasında; bir kişinin ya da bir grup insanın davranışlarını çarpıtmak ya da bilinçli şekilde karar almalarının önüne geçmek amacıyla bilinçaltı, manipülatif ya da aldatıcı teknikler kullanarak ciddi zarara sebebiyet veren sistemler ile bireylerin yaş, engellilik veya sosyal ya da ekonomik durum nedeniyle oluşan dezavantajlarından yararlanarak ciddi zarara yol açan mekanizmalar yer almaktadır. Veri koruma hukuku ve ayrımcılık yasağı bağlamında ise kişilerin ırk, siyasi görüş, sendika üyeliği, dini veya felsefi inanç, cinsel yaşam veya cinsel yönelim gibi hassas verilerine yönelik çıkarımlar yapan biyometrik sınıflandırma sistemleri ile internetten ya da kapalı devre televizyon görüntülerinden amaçsız biçimde veri kazıma yoluyla yüz tanımaya yönelik veri tabanı oluşturan ya da bu veri tabanlarını genişleten uygulamalar kesin olarak menedilmiştir. İşe alım süreçlerini de çok yakından ilgilendirecek şekilde, tıbbi ya da güvenlik gerekçeleriyle yapılan zorunlu istisnalar hariç olmak üzere, işyerlerinde ya da eğitim kurumlarında bireylerin duygularına dair çıkarımlar yapan duygu analizi sistemlerinin kullanımı da bu yasak kapsamında kabul edilemez risk olarak nitelendirilmiştir.

Tüzük'ün 6. maddesinde düzenlenen "yüksek risk" kategorisi ise kamu sağlığı, güvenliği ya da insan hakları üzerinde ciddi bir olumsuz etki yaratma potansiyeli olan yapay zekâ sistemlerini kapsamaktadır. Bu bağlamda, gerçek kişiler üzerinden profilleme yapan her türlü yapay zekâ sistemi istisnasız olarak yüksek riskli kabul edilmiştir. İş hukuku ve insan kaynakları yönetimi uygulamaları açısından bakıldığında; işe alım ve işçi seçimi için kullanılan yapay zekâ sistemleri, özellikle hedef kitle belirlenerek (algoritmik profilleme yoluyla) yayımlanan iş ilanları, başvuruların otomatik olarak analiz edilmesi, filtrelenmesi ve adayların otonom olarak değerlendirilmesi için kullanılan tüm yazılımlar doğrudan yüksek riskli sistemler sınıfına dahil edilmiştir. Kanun koyucu, bu yüksek riskli sistemleri geliştiren, kullanan, ithal eden veya kullanıma açan gerçek veya tüzel kişiler için çok katı ve şeffaf uygunluk yükümlülükleri öngörmüştür. Bu yükümlülüklerin en temel direği ise otonom kararların insani denetimini sağlamak üzere mutlak bir "insan gözetim sisteminin" kurulması zorunluluğudur.

Sürecin daha az yapısal tehlike barındıran boyutunda ise Tüzük m. 50 ve devamı maddelerinde düzenlenen "sınırlı risk" kategorisi yer almaktadır. Bu gruptaki sistemler için temel yasal kriter şeffaflık ilkesidir. Sınırlı risk teşkil eden uygulamalarda, üretilen veya işlenen ilgili içeriğin yapay zekâ tarafından oluşturulduğuna ilişkin açık ve anlaşılır bir ifadenin bulunması yasal bir zorunluluktur. Bu düzenleme, yapay zekâ ile etkileşimde bulunan insanların otonom bir sistemle muhatap oldukları konusunda net bir biçimde bilgilendirilmesini ve böylelikle dijital algoritmalara yönelik toplumsal güvenin artırılmasını amaçlamaktadır. Son olarak, tüzük kapsamında yasaklı, yüksek veya sınırlı riskli olarak sınıflandırılmayan tüm diğer yapay zekâ uygulamaları "minimum risk" kategorisinde değerlendirilmiştir. Bu kategori için tüzük bünyesinde ayrı ve spesifik yükümlülükler öngörülmemiş olup, video oyunları veya basit spam filtreleri gibi bu kapsama giren uygulamaların ortaya çıkarabileceği olası uyuşmazlıkların, Genel Veri Koruma Tüzüğü (GDPR) gibi yürürlükteki diğer genel ve sektörel mevzuatlar ile çözüme kavuşturulması öngörülmüştür.

V. AVRUPA BİRLİĞİ PLATFORM ÇALIŞMASI DİREKTİFİ

Avrupa Birliği'nin dijital emek piyasasını ve güvencesiz çalışma biçimlerini düzenlemeye yönelik en stratejik yasal hamlelerinden birini oluşturan AB Platform Çalışması Direktifi (*Platform Work Directive*)[15], iş yerlerinde algoritmik yönetim ve otomatik karar süreçlerine ilişkin işçi lehine özel ve somut güvenceler getirmektedir. Direktif kapsamında çalışanların temel haklarını korumak adına; otonom sistemler tarafından alınan otomatik kararlara karşı mutlak bir insan incelemesi talep etme hakkı, kararı üreten algoritmanın arkasındaki mantık ile süreçte kullanılan kişisel veriler hakkında detaylı bir biçimde bilgilendirilme hakkı, ön yargılı veri modellerinden kaynaklanan algoritmik ayrımcılığa karşı korunma hakkı ve nihayetinde otomasyon kaynaklı ortaya çıkan olumsuz sonuçlara esastan itiraz etme hakkı yasal güvenceye kavuşturulmuştur. Söz konusu direktif metni teknik olarak doğrudan doğruya işe alım aşamasını düzenlemeyi hedeflemese de, iş hukuku doktrininde ve kurumsal pratiklerde özellikle işe alım esnasında başvurulan otomatik eleme ve tahmine dayalı puanlama sistemlerinin yasal sınırlarının tayin edilmesi için son derece güçlü bir emsal ve yorum aracı oluşturmaktadır. Nitekim Avrupa Birliği'nin bu direktifle ortaya koyduğu genel yaklaşım; dijitalleşen iş gücü piyasasında otonom teknolojilerin insan onurunun önüne geçmesini engelleyen "insan merkezlilik", kararların arkasındaki mekanizmaları görünür kılan "şeffaflık" ve algoritmik sistemlerin hukuki denetimine imkân tanıyan "denetlenebilirlik" ilkeleri sacayağı üzerine inşa edilmiştir.

VI. TÜRKİYE'DEKİ HUKUKİ DÜZENLEMELER

Türkiye'nin yapay zekâ teknolojilerine yönelik vizyoner yaklaşımı, teorik ilkelerin ötesine geçerek somut uygulama takvimlerini barındıran Ulusal Yapay Zekâ Stratejisi 2024-2025 Eylem Planı[27] ile operasyonel bir boyut kazanmıştır. Söz konusu eylem planı, temelde ülkemizin yapay zekâ alanındaki ulusal stratejisini uluslararası normlarla tam uyumlu hale getirme ve bu ileri teknolojinin gelişimini hızlandırma amacını taşımaktadır. Yapay zekâ uzmanlarının yetiştirilmesi, girişimcilik ve yenilikçiliğin desteklenmesi, ulusal ve uluslararası iş birliklerinin güçlendirilmesi gibi birçok makro stratejik hedef içeren bu plan; aynı zamanda yapay zekâ ürünlerinin risk yönetimi, veri güvenliği ve fikri mülkiyet haklarının korunması gibi doktrinde yoğun tartışmalara konu olan alanları da kapsamaktadır. Ulusal Yapay Zekâ Stratejisi 2024-2025 Eylem Planı kapsamında öncelikli olarak altı adet ana strateji belirlenmiş olup, bunlar sırasıyla; *Yapay Zekâ Uzmanlarını Yetiştirmek ve Alanda İstihdamı Artırmak*, *Araştırma, Girişimcilik ve Yenilikçiliği Desteklemek*, *Kaliteli Veriye ve Teknik Altyapıya Erişim İmkânlarını Genişletmek*, *Sosyoekonomik Uyumu Hızlandıracak Düzenlemeleri Yapmak*, *Uluslararası Düzeyde İş Birliklerini Güçlendirmek* ve *Yapısal ve İş Gücü Dönüşümünü Hızlandırmak* şeklinde altı ana başlık altında formüle edilmiştir. Bu hedefler, özellikle iş gücü piyasasının ve insan kaynakları süreçlerinin dijitalleşme dalgasına sosyoekonomik açıdan adil bir şekilde entegre edilmesini amaçlamaktadır.

Ulusal düzeydeki bu stratejik adımlar, devletin makroekonomik ve hukuki planlama metinlerinde de güçlü bir şekilde karşılık bulmaktadır. Türkiye Cumhuriyeti Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı tarafından yakın zamanda yayımlanan ve ülkenin ekonomik/hukuki yol haritasını çizen 2025-2027 yıllarını kapsayan Orta Vadeli Program (OVP)[28], ulusal mevzuatın Avrupa Birliği Yapay Zekâ Yasası ile uyumlaştırılması için gerekli yasal düzenlemelerin kararlılıkla yapılacağına açıkça dikkat çekmektedir. Bu durum, Türkiye'nin yakın gelecekte AB müktesebatına paralel, emredici ve risk tabanlı bir yapay zekâ kanununu yürürlüğe koyacağı yönünde bir gösterge olarak değerlendirilebilecektir. Tüm bu gelişmelere ek olarak, Türkiye'deki kişisel veri koruma otoritesi olan Kişisel Verileri Koruma Kurumu (KVKK), dijitalleşmenin ve otonom sistemlerin yarattığı riskleri minimize etmek adına proaktif bir tutum sergilemiştir. Kurum, çeşitli sektörlerde yapay zekâ sistemlerinin kullanımı esnasında uyulması gereken temel insan hakları kriterlerini, veri kullanımına yönelik yasal sınırlamaları ve algoritmik şeffaflığı vurgulayan sektörel kılavuzlar yayımlamıştır. Teknik olarak doğrudan bağlayıcı hukuk kuralı niteliğinde olmayan bu rehber ve kılavuzlar; KVKK'nın, yapay zekâ tabanlı sistemleri kişisel verilerin gizliliği, veri minimizasyonu ve dürüstlük kuralı standartlarıyla entegre etme yönündeki kararlı ve dinamik denetim yaklaşımını açıkça ortaya koymaktadır.

SONUÇ

Teknoloji ve iş hukukunun kesişim noktasında yer alan işe alım süreçlerinde yapay zekâ kullanımı, dijital dönüşümün kaçınılmaz bir sonucu olarak karşımıza çıkmaktadır. İşe alım süreçlerinde ve insan kaynakları yönetiminde yapay zekâ teknolojilerinin kullanımı, operasyonel verimlilik ve maliyet yönetimi açısından avantajlar sunsa da, bireylerin temel hakları ve veri güvenliği üzerinde yönetilmesi zorunlu yapısal riskler barındırmaktadır. Kurumsal yapılara sunduğu bu avantajlar otonom sistemlerin yerini her geçen gün tahkim etse de, süreçlerin tamamen algoritmalara devredilmesi, madalyonun diğer yüzünde adayların temel hak ve özgürlükleri ile veri güvenliği üzerinde yönetilmesi zorunlu, asimetrik risk alanları doğurmaktadır. Bu doğrultuda, yapay zekâ uygulamalarının geliştirilmesi ve kurumsal pratiklerde hayata geçirilmesi sürecinde, ilgili kişilerin temel hak ve özgürlüklerine mutlak surette saygı gösterilmeli ve hiçbir hukuki hak ihlaline meydan verilmemelidir. Dijitalleşen iş gücü piyasasında otonom teknolojilerin bireyi bir nesneye indirgemesini önlemek adına; uluslararası sözleşmelerle ve anayasal metinlerle güvence altına alınan insan hakları ile temel özgürlüklerin himayesi ve

[15] Platform Work Directive, Directive (EU) 2024/2831 of the European Parliament and of the Council on improving working conditions in platform work, Official Journal of the European Union, Nov. 11, 2024.

temel insan onurunun korunması hakkı kararlılıkla gözetilmelidir.

Yapay zekâ sistemlerinin eğitildiği tarihsel veri setlerindeki ön yargıların bir sonucu olarak karşımıza çıkan cinsiyet, yaş, etnik köken ve engellilik temelli algoritmik ayrımcılık riskleri, iş gücü piyasasında fırsat eşitliği ve ayrımcılık yasağı ilkelerini derinden zedelemektedir. Bununla birlikte, adayların dijital ayak izlerinin ve sosyal medya etkileşimlerinin orantısız bir biçimde izlenerek davranışsal profil çıkarılması veya mülakatlarda otonom duygu analizlerine başvurulması, "tam gözetim" tehlikesini doğurmakta ve özel hayatın gizliliği hakkı üzerinde ağır bir baskı yaratmaktadır. Derin öğrenme modellerinin yapısal bir çıkmazı olan "kara kutu" problemi ise algoritmik kararların arkasındaki mantığın açıklanmasını imkânsız kılarak, adayların hak arama özgürlüklerinin önünde yapısal bir bariyer oluşturmaktadır.

Bu çok boyutlu riskler karşısında kişisel veri işleme temelli yapay zekâ modelleri ve veri toplama çalışmaları; kişilerin temel haklarını proaktif bir biçimde koruyan evrensel bir yaklaşım içerisinde yürütülmelidir. Bu bağlamda algoritmik süreçler; hukuka uygunluk, dürüstlük kuralı, ölçülülük, hesap verebilirlik, şeffaflık, kişisel verilerin doğru ve güncel tutulması ile kişisel veri kullanım amacının belirli, açık ve sınırlı olması ilkelerine sıkı sıkıya bağlı kalmalı ve bütüncül bir veri güvenliği yaklaşımına dayandırılmalıdır. Dolayısıyla, kişisel verilerin otonom sistemler tarafından işlenmesinde; yapay zekânın üretebileceği potansiyel risklerin önceden önlenmesi ve azaltılması üzerine odaklanan, insan haklarını, demokrasinin işleyişini, sosyal ve etik değerleri de göz önünde bulunduran çok boyutlu bir bakış açısının benimsenmesi yasal bir zorunluluktur.

Nitekim küresel hukuk düzenlerinin geliştirdiği koruma kalkanları incelendiğinde de bu risk temelli ve insan merkezli regülasyon modelinin ağırlık kazandığı görülmektedir. Avrupa Birliği; GDPR'ın getirdiği otomatik karar verme sınırlandırmalarının ardından, işe alım algoritmalarını doğrudan "yüksek riskli" kategoride konumlandıran AB Yapay Zekâ Tüzüğü ve Platform Çalışması Direktifi ile evrensel standartları belirlemiştir. Türk mevzuatı ise müstakil bir yapay zekâ kanununa henüz sahip olmamakla birlikte; 6698 sayılı KVKK'nın emredici hükümleri, Ulusal Yapay Zekâ Stratejisi Eylem Planı ve mevzuatın AB müktesebatıyla uyumlaştırılacağını ilan eden Orta Vadeli Program (OVP) doğrultusunda dinamik bir dönüşümün eşiğindedir.

Veri işleme faaliyetinin bireyler ve toplum üzerindeki sosyoekonomik etkileri bütüncül olarak değerlendirildiğinde, ortaya çıkan algoritmik çıktıların ilgili kişi (aday veya işçi) açısından her zaman kontrolü mümkün ve denetlenebilir bir yapıda tasarlanması gerekmektedir. Şayet kişisel veri işleme esaslı yapay zekâ çalışmalarında, kişisel verilerin korunması, ayrımcılık riski veya özel hayatın gizliliği açısından yüksek bir risk öngörülüyorsa, süreçlere mutlak surette bir "mahremiyet etki değerlendirmesi" uygulanmalı ve veri işleme faaliyetinin hukuka uygunluğuna ancak bu analitik çerçevede karar verilmelidir.

Netice itibarıyla, kişisel veri işleme esaslı yapay zekâ çalışmalarında, yazılımın tasarım aşamasındaki ilk adımdan itibaren kişisel verilerin korunması mevzuatına tam uyum sağlanmalı; dijital emek piyasasının geleceğinde adaleti tesis etmenin formülü, ayrımcılık yasağına sıkı sıkıya uymak, kişisel verilerin korunmasını ciddiyetle ele almak, insan denetimini sistemin merkezinde tutmak ve yüksek şeffaflık standartlarını tavizsiz bir şekilde sağlamak olacaktır.

KAYNAKÇA

[1] J. Bullock, "Artificial Intelligence, Discretion, and Bureaucracy" The American Review of Public Administration, c. 49, s. 7, 2019, ss. 1-11.

[2] M.A Brundage ve H. Anderson, "The Malicious Use Of Artificial İntelligence: Forecasting, Prevention, and Mitigation", 2018, ss. 1-100.

[3] OECD, "Recommendation of the Council on Artificial Intelligence". 2019. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449 [07.05.2026]

[4] T. Barth ve E. Arnold, "Artificial Intelligence and Administrative Discretion", The American Review of Public Administration, S. 29, 1999, ss. 332- 351.

[5] Department For Science, "Innovation & Technology, Resposible AI in Recruitment Guidance", 2024.

[6] Y. Bozkurt Gümrükçüoğlu ve G. Ahter Yakacak, "Yapay Zekânın İşe Alım Süreçlerinde Kullanımı ve Algoritmik Ayrımcılık", AÜHFD, C. 72 S. 4, 2023, ss. 1701-1757.

[7] D. Çelebi Demir, "İş İlişkisinin Kurulmasında Robot İşe Alım ve Yapay Zekâ Kullanımı", DEÜHFD, c. 27, sy 1, ss. 205–253, May. 2025.

[8] Z. Chen, "Ethics and discrimination in artificial intelligence-enabled recruitment practices," Humanit. Soc. Sci. Commun., vol. 10, no. 1, Sep. 2023.

[9] Y. Çankaya, "Yapay zekanın iş ilişkisine etkileri," Doktora Tezi, Atılım Üniversitesi Sosyal Bilimler Enstitüsü, Ankara, 2024.

[10] U. Baysal, "Avrupa Birliği Yapay Zekâ Tüzüğü çerçevesinde işin ve işyerinin yönetiminde yapay zekâ sistemleri kullanımı," Sicil İş Hukuku Dergisi, no. 54, pp. 108–125, 2025.

[11] A. Güzel, D. Ugan Çatalkaya, and H. Heper, "İş hukukunun yapay zekâ ile buluşması: İşverenin algoritmik yönetimi," Hukuk ve Adalet Eleştirel Hukuk Dergisi, vol. 15, Özel Sayı, ss. 25–110, 2023.

[12] G. Alpagut and A. Karaca Yağcı, "İşyerinde yapay zekâ uygulaması ve ayrımcılık," Sicil İş Hukuku Dergisi, vol. 2, no. 49, ss. 11–44, 2023.

[13] V. Papadouli, "AI in recruiting: Equality and non-discrimination considerations in the light of the AI Act, AGG and BGB," in Artificial Intelligence in German Employment Relationships, D. Wasser and V. Papakonstantinou, Eds. Baden-Baden, Germany: Nomos, Jan. 2026, s. 234-235.

[14] A. Kelly-Lyth, "Algorithmic discrimination at work," European Labour Law Journal, vol. 14, no. 2, ss. 152–171, 2023; H. Parviainen, "Challenges of direct discrimination in algorithmic recruitment: Insuperable or not?," International Journal of Comparative Labour Law and Industrial Relations, vol. 40, no. 4, ss. 437–466, 2024.

[15] X. Chang, "Gender bias in hiring: An analysis of the impact of Amazon's recruiting algorithm," Advances in Economics, Management and Political Sciences, vol. 23, no. 1, ss. 134–140, Sep. 2023.

[16] E. Winick, "Amazon ditched AI recruitment software because it was biased against women," MIT Technology Review, Oct. 10, 2018. [Online]. Available: https://www.technologyreview.com/2018/10/10/139858/amazon-ditched-ai-recruitment-software-because-it-was-biased-against-women/ [07.05.2026].

[17] "Facebook's ad delivery algorithm is discriminating based on race, gender and age in photos, Northeastern researchers find," Northeastern Global News, Oct. 25, 2022. [Online]. Available: https://news.northeastern.edu/2022/10/25/facebook-algorithm-discrimination/ [07.05.2026]

[18] "Facebook's job ads algorithm is sexist, French equality watchdog rules," The Guardian, Nov. 5, 2025. [Online]. Available: https://www.theguardian.com/world/2025/nov/05/facebook-job-ads-algorithm-is-sexist-french-equality-watchdog-rules [07.05.2026].

[19] J. G. Asare, "What the Workday lawsuit reveals about AI bias—And how to prevent it," Forbes, Jun. 23, 2025. [Online]. Available: https://www.forbes.com/sites/janicegassam/2025/06/23/what-the-workday-lawsuit-reveals-about-ai-bias-and-how-to-prevent-it/ [07.05.2026]

[20] J. Campbell, "The Workday lawsuit just changed everything: Why your hiring process needs an urgent audit," LinkedIn, 2025. [Online]. Available: https://www.linkedin.com/pulse/workday-lawsuit-just-changed-everything-why-your-hiring-campbell-8xcye/ [07.05.2026].

[21] M. C. Aktan, "Üretken yapay zeka ile işe alım sürecinde aday işçinin kişisel verilerinin örneklerle değerlendirilmesi," Yeditepe Üniversitesi Hukuk Fakültesi Dergisi, vol. 22, no. 2, ss. 507–550, 2025.

[22] M. Bulut, "Özel bir hukuksal koruma ve veri kategorisi alanı: Hassas kişisel veriler," Ankara Barosu Dergisi, vol. 78, no. 3, ss. 99–150, 2020.

[23] M. Özer Deniz, "İşçilerin özel nitelikli kişisel verilerinin korunması ve bundan doğan sorumluluk," Türkiye Adalet Akademisi Dergisi, no. 45, ss. 355–378, 2021.

[24] İ. Urhanoğlu and D. Kaynar, "İşçinin kişisel verilerinin korunması kapsamında unutulma hakkı," in Hasan Tahsin Fendoğlu'na Armağan. Ankara, Türkiye: Yetkin Yayınları, 2023, ss. 505–557.

[25] A. Aloisi and E. Gramano, "Artificial intelligence is watching you at work: Digital surveillance, employee monitoring, and regulatory issues in the EU context," Comparative Labor Law & Policy Journal, vol. 41, no. 1, ss. 95–120, Fall 2019.

[26] I. Ajunwa, "The 'black box' at work," Big Data & Society, vol. 7, no. 2, 2020, Art. no. 2053951720966181; V. V. Fomin and P. Astromskis, "The black box problem," in Future Law, Ethics, and Smart Technologies, 2023.

[27] T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi Başkanlığı ve Sanayi ve Teknoloji Bakanlığı, Ulusal Yapay Zekâ Stratejisi 2024-2025 Eylem Planı, Ankara, Türkiye, 2024. [Online]. Available: https://www.sanayi.gov.tr/assets/pdf/UlusalYapayZekaStratejisi2024-2025EylemPlani.pdf [07.05.2026]

[28] Cumhurbaşkanlığı, Orta Vadeli Program (2025-2027), Resmî Gazete, no. 32653 (1. Mükerrer), 5 Eylül 2024.

İnternet Kaynakları

https://doi.org

https://legalinstruments.oecd.org

https://news.northeastern.edu

https://textio.com/

https://www.ilo.org

https://www.oecd.org

https://www.reuters.com

https://www.sanayi.gov.tr/

https://www.taledo.com/de/

https://www.weforum.org

Adana ilindeki kavşak tasarım uygulamalarının mikrosimülasyon destekli optimizasyonu

1st Gülşah Doğan
Akıllı Ulaşım Sistemleri ve Teknolojileri
Bandırma Onyedi Eylül Üniversitesi
Balıkesir, Türkiye
gulsahmuslubas@gmail.com
0009-0007-3403-3766

2nd Dr. Seyitali İlyas
Elektrik ve Elektronik Mühendisliği
Kırgızistan – Türkiye Manas Üniversitesi
Bişkek, Kırgızistan
seyitali.ilyas@manas.edu.kg
0000-0003-2009-3454

3rd Prof.Dr.Mehmet Tektaş
Akıllı Ulaşım Sistemleri ve Teknolojileri
Bandırma Onyedi Eylül Üniversitesi
Balıkesir, Türkiye
mtektas@bandirma.edu.tr
0000-0001-9564-8069

4th Prof.Dr.Necla Tektaş
Akıllı Ulaşım Sistemleri ve Teknolojileri
Bandırma Onyedi Eylül Üniversitesi
ntektas@bandirma.edu.tr
0000-0002-8190-4532

***Özet*—Bu çalışma, kentsel ulaşım sistemlerinde kavşak tasarımının trafik işletimi, güvenlik ve sürdürülebilirlik üzerindeki etkilerini ortaya koymak amacıyla gerçekleştirilmiştir. Çalışma kapsamında, Adana ili Seyhan ilçesinde yer alan iki önemli kavşakta uygulanan tasarım ve düzenleme örnekleri incelenmiştir. İlk olarak, mevcut hemzemin kavşak düzeni ile üst geçitli *Single Point Urban Interchange* (SPUI) tasarımı, PTV VISSIM mikrosimülasyon yazılımı kullanılarak karşılaştırmalı biçimde modellenmiş ve performans açısından değerlendirilmiştir. İkinci olarak, Ali Bozdoğanoğlu Bulvarı ile Aliya İzzet Begoviç Bulvarı kesişiminde yer alan dönel adalı kavşakta gözlenen trafik tıkanıklığı ve güvenlik sorunlarının giderilmesine yönelik gerçekleştirilen fiziksel düzenleme ele alınmıştır. Analiz sonuçları, SPUI tasarımının kavşak hizmet düzeyini LOS F'den LOS D'ye yükselttiğini, ortalama gecikme süresini 166 saniyeden 54 saniyeye düşürdüğünü, ortalama araç hızını 10 km/saatten 25 km/saate çıkardığını ve yakıt tüketiminde %44 oranında iyileşme sağladığını ortaya koymuştur. Ayrıca, dönel adalı kavşakta yapılan fiziksel düzenleme sonrasında kaza verilerinde belirgin bir azalma olduğu belirlenmiştir. Elde edilen bulgular, mikrosimülasyon modellemesinin kentsel kavşak tasarım ve iyileştirme süreçlerinde karar destek aracı olarak önemli bir işleve sahip olduğunu; ekonomik, operasyonel ve güvenlik açısından etkin sonuçlar sunduğunu göstermektedir.**

Anahtar Kelime—kavşak tasarımı, mikrosimülasyon, VISSIM, SPUI, hizmet düzeyi, trafik güvenliği, TGAP, Adana

***Abstract*—This study was conducted to reveal the effects of intersection design on traffic operation, safety, and sustainability in urban transportation systems. Within the scope of the study, design and improvement applications implemented at two major intersections located in the Seyhan district of Adana were examined. First, the existing at-grade intersection layout and an overpass-based Single Point Urban Interchange (SPUI) design were comparatively modeled using the PTV VISSIM microsimulation software and evaluated in terms of performance. Second, a physical improvement implemented to address traffic congestion and safety problems observed at the roundabout located at the intersection of Ali Bozdoğanoğlu Boulevard and Aliya İzzet Begovic Boulevard was analyzed. The results indicated that the SPUI design improved the level of service of the intersection from LOS F to LOS D, reduced the average delay from 166 seconds to 54 seconds, increased the average vehicle speed from 10 km/h to 25 km/h, and achieved a 44% improvement in fuel consumption. In addition, a significant reduction in crash data was identified following the physical improvement carried out at the roundabout.**

The findings demonstrate that microsimulation modeling serves as an important decision-support tool in urban intersection design and improvement processes, providing effective outcomes in economic, operational, and safety-related terms.

Keywords: intersection design, microsimulation, VISSIM, SPUI, level of service, traffic safety, TSAP,Adana

I. Giriş

Kentsel nüfus artışı ve motorlu taşıt sahipliğindeki hızlı yükseliş, dünya genelinde kent merkezlerindeki yol altyapısının kapasitesini önemli ölçüde aşmaktadır. Kavşaklar, birden fazla yolun kesiştiği ve farklı yönlerdeki trafik akımlarının bir araya geldiği bu yapılar, kentsel ulaşım ağının en karmaşık ve en kritik bileşenleri olarak öne çıkmaktadır. Trafik yoğunluğunun kavşaklarda neden olduğu gecikme, yakıt tüketimi artışı, egzoz gazı emisyonu ve trafik kazaları hem bireysel sürücüler hem de toplumsal ekonomi üzerinde ciddi olumsuz etkiler doğurmaktadır.

Araştırmalar, kentsel yollardaki toplam tıkanıklığın yaklaşık %40'ının kavşak noktalarından kaynaklandığını ve bu durumun milyarlarca litre yıllık fazla yakıt tüketimine yol açtığını göstermektedir [1]. Konuyla ilgili yapılan çalışmada, trafik tıkanıklığından kaynaklanan gecikmelerin sürücülerin yıllık 7 milyar saatten fazla ekstra süre harcamasına ve 3 milyar galon yakıt israfına neden olduğu tespit edilmiştir [2].

Öte yandan kavşaklardaki yetersiz ya da hatalı tasarımlar trafik kazalarının başlıca tetikleyicileri arasında yer almakta; yaralanmalı ve ölümlü kazaların büyük bölümü bu noktalarda yoğunlaşmaktadır [3].

Bilgisayar teknolojilerindeki gelişmeler, kavşak tasarımının değerlendirilmesinde mikrosimülasyon modellemesini giderek öne çıkarmıştır. PTV VISSIM gibi araçlar, trafik akışını bireysel araç düzeyinde simüle ederek gerçek koşulları yüksek doğrulukla yansıtabilmekte; böylece mühendislere kavşakları inşa etmeden önce farklı tasarım senaryolarını karşılaştırma imkânı sunmaktadır [4]. Nitekim VISSIM tabanlı çalışmalar, gerçek trafik parametreleriyle kalibre edilmiş modellerin sahadan elde edilen verilere %10'un altında hata payıyla yakınsadığını ortaya koymaktadır [5].

Hızla büyüyen Adana kenti, bu tür tasarım sorunlarının somut biçimde gözlemlendiği bir örnek oluşturmaktadır. Özellikle otoyol bağlantı güzergahları üzerindeki kavşaklar, artan araç yoğunluğu, yetersiz kapasite ve düşen hizmet seviyeleri nedeniyle acil müdahale gerektiren sorunlu noktalar haline gelmiştir. Yapılan sayım çalışmaları, söz konusu kavşaklarda araç yoğunluğunun eski verilere kıyasla önemli bir artış kaydettiğini göstermiştir. Bu çalışma, Adana ili Seyhan ilçesinde iki kritik kavşakta yürütülen trafik çalışmaları sunmaktadır. Araştırmanın özgün katkısı; hem Single Point Urban Interchange (SPUI) tasarımının hem de dönel adalı kavşak fiziksel düzenlemesinin mikrosimülasyon destekli analizini, gerçek kaza verileri ve ekonomik analiz bulgularıyla birlikte bütüncül bir çerçevede ele alınmasından kaynaklanmaktadır.

I. MATERYAL VE YÖNTEM

Bu bölümde araştırmanın çalışma alanı, veri toplama yöntemleri, kullanılan simülasyon yazılımı ve analiz yaklaşımı ayrıntılı biçimde açıklanmaktadır.

Bu çalışma kapsamında, Adana ili Seyhan ilçesinde yer alan iki farklı kavşak alanı inceleme konusu olarak seçilmiştir. İlk çalışma alanı, Öğretmenler Bulvarı ile Aliya İzzet Begoviç Bulvarı'nın kesişim noktasında yer almaktadır. Söz konusu kavşak, Adana Çevre Yolu bağlantısını sağlayarak otoyol erişiminde önemli bir arteriyel düğüm noktası işlevi görmektedir. İkinci çalışma alanı ise yine Seyhan ilçesinde, Ali Bozdoğanoğlu Bulvarı ile Aliya İzzet Begoviç Bulvarı'nın kesişiminde bulunan ve dönel ada düzenine sahip kavşağı kapsamaktadır. Bu kavşak, mevcut geometrik ve işletimsel özellikleri bakımından çalışma kapsamında ayrıca değerlendirilmiştir.

Her iki kavşak için kapsamlı bir saha araştırması yürütülmüştür. Araç sayım verileri, kavşakların tüm yaklaşım kollarında manuel sayım ile insansız hava araçları kullanılarak elde edilmiştir. Kaza verileri, T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı Coğrafi Bilgi Sistemleri Genel Müdürlüğü (CBSGM) tarafından geliştirilen Trafik Güvenliği Analiz Platformu (TGAP) üzerinden temin edilmiş ve her iki kavşak için üç yıllık dönem esas alınarak değerlendirilmiştir [6-7]. Kavşaklara ilişkin geometrik veriler; uydu görüntüleri, yerinde yapılan arazi ölçümleri ve Adana Büyükşehir Belediyesi'nden temin edilen kadastro planları doğrultusunda elde edilmiştir [8]. Bu çalışmada, kavşak performanslarının değerlendirilmesinde PTV Group tarafından geliştirilen VISSIM (Verkehr In Städten SIMulationsmodell) mikrosimülasyon yazılımı kullanılmıştır.

Çalışma kapsamında incelenen her iki kavşak için mevcut durumu temsil eden temel senaryo ile önerilen veya uygulanan tasarım alternatifleri ayrı ayrı modellenmiş ve karşılaştırmalı olarak değerlendirilmiştir. Böylece, mevcut işletim koşulları ile alternatif tasarım çözümlerinin kavşak performansı üzerindeki etkilerinin nicel olarak ortaya konulması amaçlanmıştır.

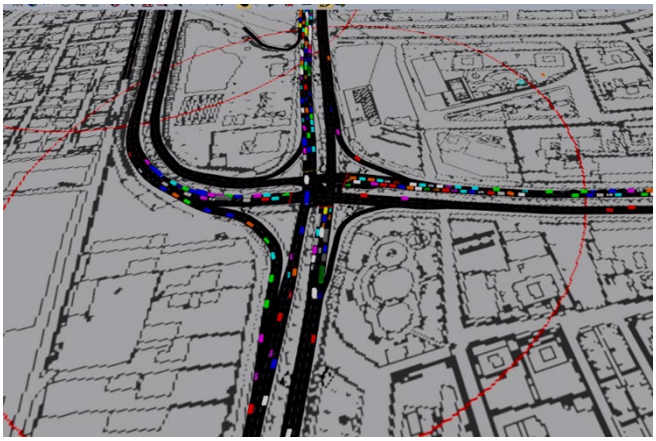
Performans değerlendirmelerinde, Highway Capacity Manual (HCM) kapsamında tanımlanan Hizmet Düzeyi (Level of Service – LOS) temel ölçütlerden biri olarak ele alınmıştır. Buna ek olarak, kavşakların işletim performansını daha ayrıntılı biçimde değerlendirebilmek amacıyla ortalama bekleme süresi (saniye/araç), toplam sefer süresi, toplam bekleme süresi ve ortalama hız (km/saat) göstergeleri kullanılmıştır [9]. Bu göstergeler aracılığıyla, farklı tasarım alternatiflerinin trafik akışı üzerindeki etkileri çok boyutlu olarak analiz edilmiştir.

Ekonomik değerlendirme aşamasında ise simülasyon modelinden elde edilen yakıt tüketimi ve seyahat süresi verileri esas alınmıştır. Bu kapsamda, yalnızca mevcut trafik koşulları değil, aynı zamanda gelecekte oluşabilecek talep artışını temsil etmek amacıyla araç yoğunlukları %15 oranında artırılarak gelecek yıl senaryosu da oluşturulmuştur. Ayrıca, kavşakların işletim verimliliğini artırmaya yönelik olarak en uygun sinyal süresi ve faz planı, VISSIM optimizasyon modülü kullanılarak belirlenmiştir. Bu yaklaşım sayesinde, önerilen tasarımların hem mevcut koşullar hem de artan trafik talebi altında göstereceği performans bütüncül bir çerçevede değerlendirilmiştir.

II. BULGULAR

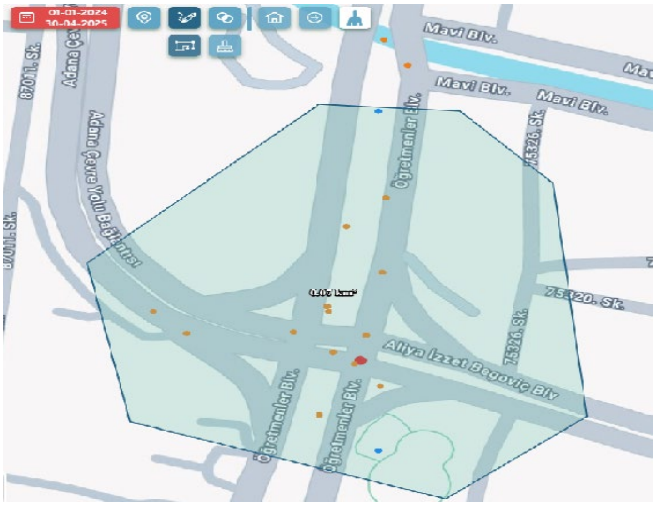
A. Öğretmenler Bulvarı ile Aliya İzzet Begoviç Bulvarı kesişim kavşağı

Adana ili Seyhan ilçesinde, Öğretmenler Bulvarı ile Aliya İzzet Begoviç Bulvarı kesişimindeki dört kollu hemzemin kavşakta 12.10.2021 tarihinde yapılan araç sayımları, sabah ve akşam zirve saatlerde yüksek trafik talebi ve yoğun taşıt hareketleri olduğunu göstermiştir; geçmiş yıllarla karşılaştırıldığında tüm yönlerde yaklaşık %15 oranında trafik artışı tespit edilmiş, bu durum kavşağın mevcut kapasitesini zorlayarak gecikme ve kuyruklanmaları artırmıştır. Elde edilen veriler, mevcut işletim koşullarının yetersiz kaldığını ortaya koymuş ve kavşağın performansını iyileştirmek amacıyla alternatif tasarım çözümlerinin değerlendirilmesi gerekliliğini ortaya çıkarmıştır.



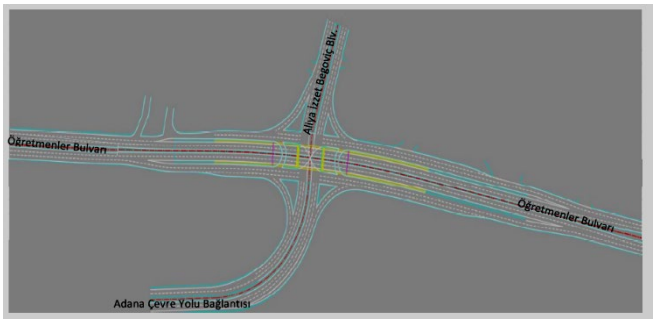
Şekil 1. Öğretmenler Bulvarı – Aliya İzzet Begoviç Bulvarı mevcut kavşak düzeninin PTV VISSIM ortamında oluşturulan mikrosimülasyon modeli

Şekil 1'de gösterilen PTV VISSIM tabanlı mikrosimülasyon modeli, Öğretmenler Bulvarı ile Aliya İzzet Begoviç Bulvarı kesişimindeki dört kollu hemzemin kavşağın mevcut geometrisi, trafik hacimleri ve işletim koşulları dikkate alınarak oluşturulmuş; analizler özellikle zirve saatlerde yoğun taşıt birikimi ve belirgin kuyruklanmalar olduğunu ortaya koymuştur. Simülasyon sonuçları gecikme, bekleme süresi, hız ve hizmet düzeyi gibi performans göstergelerini nicel olarak belirlemiş ve mevcut kavşak düzeninin artan trafik talebini karşılamada yetersiz kaldığını, bu nedenle alternatif tasarım çözümlerinin değerlendirilmesi gerektiğini göstermiştir.



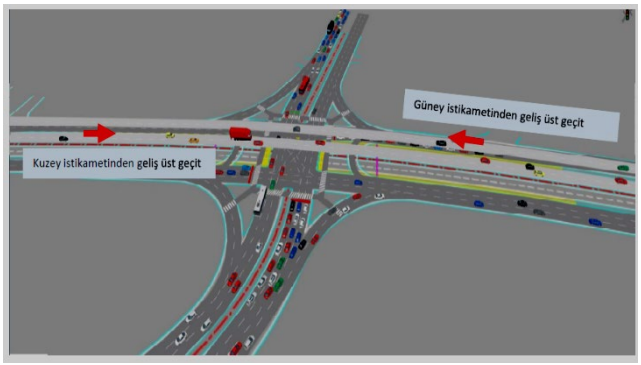
Şekil 2. İncelenen kavşakta 01.01.2024 – 30.04.2025 tarihlerindeki döneme ait trafik kazalarının mekânsal dağılımı ve zamansal yoğunluk haritası

Şekil 2'de sunulan veriler, Öğretmenler Bulvarı ile Aliya İzzet Begoviç Bulvarı kesişiminde 16 aylık dönemde meydana gelen 38 kazanın mekânsal ve zamansal dağılımını ortaya koymakta; kazaların özellikle kavşak giriş-çıkışları ve çatışma noktalarında yoğunlaştığını göstermektedir. Kazaların 2'si ölümlü, 33'ü yaralanmalı ve 2'si maddi hasarlı olup, bu durum kavşağın hem işletim hem de trafik güvenliği açısından risk taşıdığını ortaya koymaktadır. Zamansal analizler, kazaların çoğunlukla zirve saatlerde, hafta içi günlerde ve açık hava koşullarında meydana geldiğini, dolayısıyla temel nedenin çevresel faktörlerden ziyade yüksek trafik hacmi, geometrik yapı ve işletim koşulları olduğunu işaret etmektedir. Bu bulgular doğrultusunda, mevcut dört kollu hemzemin kavşak yerine önerilen üst geçitli SPUI tasarımı, %15 artırılmış trafik talebi altında mikrosimülasyon ile analiz edilmiş ve uygun sinyalizasyon düzeni belirlenerek gelecekteki performansı değerlendirilmiştir.



Şekil 3. Single point kavşak tasarımı

Elde edilen bulgular, önerilen üst geçitli SPUI tasarımının mevcut hemzemin kavşağa kıyasla daha etkin bir trafik işletimi sağladığını göstermektedir. Ana akımların farklı kotlarda ayrılmasıyla çatışma noktaları azaltılmış, taşıt hareketleri daha düzenli hale getirilmiş ve özellikle kuzey-güney yönündeki bağlantıların kavşak yükünü hafiflettiği görülmüştür. Mikrosimülasyon sonuçları, kuyruklanma ve gecikmelerin azaldığını, trafik akışının daha süreklilik kazandığını ve kapasite kullanımının iyileştiğini ortaya koymuştur. Bu doğrultuda, SPUI tasarımının yalnızca mevcut sorunları gidermekle kalmayıp, gelecekte artacak trafik talebine karşı da daha sürdürülebilir ve verimli bir çözüm sunduğu sonucuna ulaşılmıştır.



Şekil 4. Simülasyon program analizi

Şekil 4 ve Şekil 5'te sunulan uygulama sonrası görüntüler, Öğretmenler Bulvarı ile Aliya İzzet Begoviç Bulvarı kesişiminde gerçekleştirilen köprülü kavşak projesinin yeni geometrik düzenini ortaya koymakta; üst geçit, bağlantı kolları ve kanalize edilmiş trafik akımları açık biçimde görülmektedir. İHA görüntüsü, taşıt hareketlerinin farklı kotlarda ayrıştırıldığını ve özellikle ana arterde kesintisiz akımın sağlandığını gösterirken, bu düzenleme sayesinde kavşak içindeki çatışma noktalarının azaltıldığı ve trafik akışının daha düzenli hale getirildiği anlaşılmaktadır.



Şekil 5. Uygulanan köprülü kavşak projesinin insansız hava aracı görüntüsü

Şekil 6'da ise uygulanan köprülü kavşağın saha düzeyindeki görünümü sunulmaktadır. Bu görüntüde, alt kotta işletilen sinyalize kavşak düzeni ile üst kotta devam eden ana trafik akımı birlikte izlenebilmektedir.

Böylece proje sonrasında kavşakta hem ana yön hareketlerinin sürekliliğinin sağlandığı hem de alt kotta kalan dönüş ve dağılım hareketlerinin kontrollü biçimde yönetildiği anlaşılmaktadır. Sinyalizasyon elemanları, şerit düzenlemeleri, yaya geçitleri ve kanalizasyon adaları da kavşağın yalnızca kapasite artışı amacıyla değil, aynı zamanda trafik güvenliği ve işletim düzeninin iyileştirilmesi amacıyla tasarlandığını göstermektedir.



Şekil 6. Uygulanan köprülü kavşak projesinin insansız hava aracı görüntüsü

Uygulama sonrası elde edilen fiziksel düzenlemenin, mikrosimülasyon bulgularıyla uyumlu olduğu görülmüş; köprülü kavşak çözümünün taşıt akımlarını daha düzenli hale getirerek bekleme ve gecikme sürelerini azalttığı ve kavşak kapasitesini artırdığı belirlenmiştir. Ana arter trafiğinin alt kot hareketlerinden ayrılmasıyla özellikle yoğun saatlerdeki sıkışıklığın önemli ölçüde azaldığı değerlendirilmekte olup, bu proje kavşak tasarım optimizasyonunun sahadaki başarılı bir uygulaması olarak öne çıkmaktadır.



Şekil 7. Uygulama sonrası kavşak alanına ait kaza yoğunluk haritası

Şekil 7'de sunulan veriler, uygulama sonrası kavşak ve çevresindeki trafik kazalarının mekânsal dağılımını ve istatistiklerini ortaya koymakta; kazaların belirli bölgelerde kümelenmesine rağmen önceki döneme kıyasla belirgin bir azalma olduğunu göstermektedir. İncelenen sekiz aylık süreçte 3 maddi hasarlı, 3 yaralanmalı kazanın meydana gelmesi ve ölümlü kazaların görülmemesi, köprülü kavşak düzenlemesinin trafik güvenliğini önemli ölçüde artırdığını ortaya koymaktadır. Ayrıca kazaların kavşak merkezi yerine daha çok bağlantı ve geçiş bölgelerinde yoğunlaşması, ana akımların farklı kotlarda ayrılmasıyla çatışma noktalarının azaldığını ve yeni geometrik düzen ile sinyalizasyonun güvenlik performansını iyileştirdiğini göstermektedir.

TABLO 1.Single Point Urban Interchange (SPUI) ve Mevcut Hemzemin Kavşak Tasarımlarının Karşılaştırmalı Performans Sonuçları

Performans parametreleri	***Eski tasarım***	***Uygulanan tasarım (SPUI)***
Hizmet düzeyi (LOS)	LOS F	LOS D
Toplam taşıt sayısı (1 saat)	6.200	10.645
Ortalama bekleme süresi (saniye)	166	54
Toplam sefer süresi (saat)	1.446.980	842.146
Toplam bekleme süresi (saat)	1.110.470	485.720
Ortalama hız (km/sa)	10	25

PTV VISSIM mikrosimülasyon analizleri, önerilen SPUI tasarımının mevcut hemzemin kavşağa kıyasla tüm temel performans göstergelerinde önemli iyileşmeler sağladığını ortaya koymuştur. Kavşağın hizmet düzeyi LOS F'ten LOS D'ye yükselirken, saatlik taşıt kapasitesi 6.200'den 10.645'e çıkmış, ortalama bekleme süresi 166 saniyeden 54 saniyeye düşmüş ve ortalama hız 10 km/sa'den 25 km/sa'ye yükselmiştir. Bu sonuçlar, yeni tasarımın daha yüksek trafik talebini karşılayabildiğini, gecikmeleri azalttığını ve trafik akışını daha verimli ve düzenli hale getirdiğini göstermektedir.

Ekonomik ve çevresel analiz

Simülasyon verilerine dayalı fayda – maliyet analizi, önerilen SPUI tasarımının hem ekonomik hem de çevresel açıdan önemli avantajlar sağladığını göstermektedir. Yakıt tüketiminde %44 oranında azalma elde edilirken, bu

tasarrufun önemli düzeyde ekonomik kazanç yarattığı belirlenmiştir. Ayrıca günlük yakıt tasarrufu sayesinde emisyonların azalması ve çevresel sürdürülebilirliğe katkı sağlanması, tasarımın etkinliğini destekleyen diğer önemli bulgulardır. Bu sonuçlar, kentsel kavşak iyileştirmelerinin verimlilik ve çevre açısından güçlü faydalar sunduğunu ortaya koymaktadır.

TABLO 2.Single Point Urban Interchange (SPUI) Tasarımına İlişkin Ekonomik ve Çevresel Analiz Sonuçları

Gösterge	***Değer***
Yakıt tasarrufu oranı	%44
Saatlik ekonomik fayda	1.864434,00 USD
Günlük ekonomik fayda	11.186.604,00 USD
Yıllık ekonomik fayda	2.684.784.960,00 USD
Günlük yakıt tasarrufu	5.148 litre
Günlük emisyon azalımı	78 kg CO_2
Eşdeğer çevresel fayda	8,1 ağaç/gün

Single point urban interchange (SPUI) tasarımı ve simülasyon sonuçları

SPUI projesinin hizmete alınmasının ardından elde edilen sekiz aylık kaza verileri, kavşakta ölümlü kazaların ortadan kalktığını ve yalnızca sınırlı sayıda maddi hasarlı kazanın meydana geldiğini göstermektedir. Proje öncesi dönemdeki yüksek kaza sayılarıyla karşılaştırıldığında, bu durum trafik güvenliği açısından belirgin bir iyileşmeye işaret etmektedir. Bulgular, SPUI tasarımının çatışma noktalarını azaltarak ve trafik akışını daha kontrollü hâle getirerek güvenliği önemli ölçüde artırdığını ortaya koymakta; aynı zamanda literatürde yer alan çalışmalarla uyumlu şekilde, bu tür kavşak düzenlemelerinin güvenlik performansı üzerinde olumlu etkiler sağladığını desteklemektedir.

B. Ali Bozdoğanoğlu Bulvarı – Aliya İzzet Begoviç Bulvarı dönel adalı hemzemin kavşağı

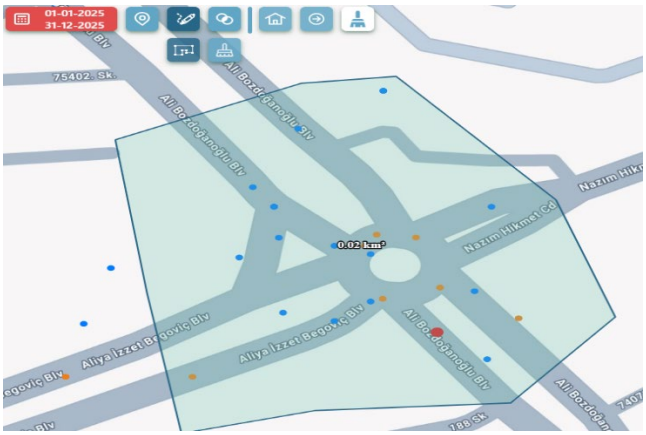
Adana ili Seyhan ilçesinde yer alan söz konusu dönel adalı hemzemin kavşakta gerçekleştirilen araç sayımları, özellikle sabah ve akşam zirve saatlerinde yüksek trafik talebi ve yoğun taşıt hareketlerinin oluştuğunu ortaya koymuştur. Kavşağın mevcut geometrik yapısı ve sınırlı depolama alanı, kentsel akış, ticari faaliyetler ve toplu taşıma bağlantılarıyla birleşerek işletim performansını olumsuz etkilemektedir. Elde edilen veriler; kapasite yetersizliği, gecikme ve kuyruklanma sorunlarının yanı sıra yaya güvenliği eksikliklerini de ortaya koymuş ve mevcut kavşak düzeninin yetersiz kaldığını göstererek alternatif tasarım çözümlerinin gerekliliğini açıkça ortaya koymuştur.



Şekil 8. Ali Bozdoğanoğlu Bulvarı – Aliya İzzet Begoviç Bulvarı mevcut kavşak düzeninin PTV VISSIM ortamında oluşturulan mikrosimülasyon modeli

Şekil 8'de sunulan PTV VISSIM mikrosimülasyon modeli, kavşağın mevcut geometrik yapısı, trafik hacimleri ve işletim koşulları dikkate alınarak oluşturulmuştur. Simülasyon sonuçları, özellikle zirve saatlerde kavşakta yoğun taşıt birikimleri, yaklaşım kollarında belirgin kuyruklanmalar ve dönel ada içerisindeki yetersiz alan nedeniyle trafik akışında aksaklıklar yaşandığını ortaya koymaktadır. Ayrıca model, gecikme, bekleme süresi, hız ve hizmet düzeyi gibi temel performans göstergelerinin nicel olarak değerlendirilmesine imkân sağlamıştır.

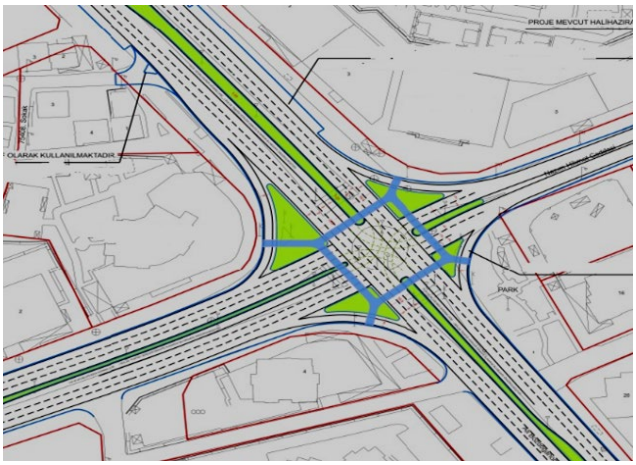
Elde edilen bulgular, mevcut dönel adalı hemzemin kavşak düzeninin artan trafik talebini karşılamada yetersiz kaldığını ve alternatif tasarım çözümlerinin gerekliliğini açıkça ortaya koymaktadır.



Şekil 9. İncelenen kavşakta 01.01.2025 – 31.12.2025 tarih aralığı dönemine ait trafik kazalarının mekânsal dağılımı ve zamansal yoğunluk haritası

Şekil 9'da sunulan veriler, Ali Bozdoğanoğlu Bulvarı ile Aliya İzzet Begoviç Bulvarı kesişiminde meydana gelen kazaların mekânsal ve zamansal dağılımını ortaya koymakta; kazaların özellikle kavşak giriş-çıkışları ve çatışma noktalarında yoğunlaştığını göstermektedir. İncelenen bir yıllık dönemde toplam 24 kaza meydana gelmiş olup, bunların 1'i ölümlü, 10'u yaralanmalı ve 13'i maddi hasarlı kazalardan oluşmaktadır. Yaralanmalı kazaların yüksek oranı, kavşakta trafik güvenliği açısından önemli riskler bulunduğuna işaret etmektedir.

Zamansal analizler, kazaların özellikle zirve saatlerde ve belirli günlerde yoğunlaştığını; hava ve aydınlatma koşullarından ziyade trafik hacmi, kavşak geometrisi ve işletim özelliklerinin kaza oluşumunda daha belirleyici olduğunu göstermektedir. Bu bulgular doğrultusunda, Ali Bozdoğanoğlu Bulvarı ile Aliya İzzet Begoviç Bulvarı kesişiminde bulunan mevcut dönel adalı kavşak yerine önerilen SPUI tasarımı, gelecekteki trafik talebi dikkate alınarak mikrosimülasyon ile değerlendirilmiştir. Bu kapsamda trafik hacimleri %15 artırılarak analiz yapılmış, artan yük altında kavşağın performansı incelenmiş ve en uygun sinyal süreleri ile faz planı belirlenmiştir.



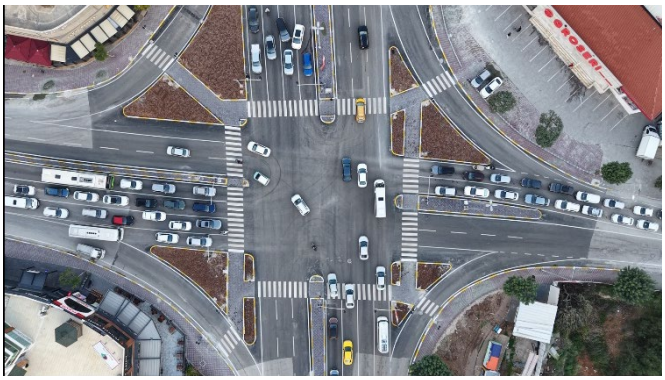
Şekil 10. Single point kavşak tasarımı

Elde edilen bulgular, önerilen SPUI tasarımının mevcut kavşak düzenine kıyasla daha etkin, düzenli ve güvenli bir trafik işletimi sağladığını ortaya koymaktadır. Çatışma noktalarının azaltılması ve dönüş hareketlerinin tek noktadan sinyalize şekilde kontrol edilmesi sayesinde trafik akışı daha akıcı ve kontrollü hâle getirilmiştir. Mikrosimülasyon sonuçları, kuyruklanma ve gecikmelerin azaldığını, akış sürekliliğinin arttığını ve kapasite kullanımının iyileştiğini göstermektedir. Bu doğrultuda, SPUI tasarımının yalnızca mevcut işletim sorunlarını gidermekle kalmadığı, aynı zamanda gelecekte artması beklenen trafik talebine karşı da sürdürülebilir ve güvenli bir çözüm sunduğu sonucuna ulaşılmıştır.



Şekil 11. Simülasyon program analizi

Şekil 11 ve Şekil 12, Ali Bozdoğanoğlu Bulvarı ile Aliya İzzet Begoviç Bulvarı kesişiminde uygulanan SPUI tasarımının proje sonrası durumunu görsel olarak ortaya koymaktadır. İnsansız hava aracı ile elde edilen üstten görünüm, kavşağın yeni geometrik düzenini, bağlantı kollarını ve kanalize edilmiş trafik akımlarını açık biçimde göstermektedir. Bu düzenleme ile taşıt hareketlerinin ayrıştırıldığı ve kavşak içindeki akım organizasyonunun daha sistematik hâle getirildiği anlaşılmaktadır. Özellikle ana arter üzerindeki kesintisiz trafik akışının sağlanması, çatışma noktalarının azaltılmasına yönelik mühendislik yaklaşımının sahaya başarılı şekilde yansıtıldığını göstermektedir.



Şekil 12. Uygulanan single point kavşak projesinin insansız hava aracı görüntüsü

Şekil 12'de sunulan saha görünümü, uygulanan kavşak tasarımının pratikteki etkilerini açık biçimde ortaya koymaktadır. Tüm hareketlerin tek bir kavşak üzerinde entegre edilmesiyle çatışma noktalarının minimize edildiği; geçiş öncelikleri, koridor düzeni ve yaya güvenliğinin optimize edildiği anlaşılmaktadır. Ayrıca sinyalizasyon elemanları, şerit düzenlemeleri, yaya geçitleri ve kanalizasyon adaları, tasarımın yalnızca kapasite artışına değil, aynı zamanda trafik güvenliği ve işletim düzeninin iyileştirilmesine odaklandığını göstermektedir.

Uygulama sonrası elde edilen fiziksel düzenleme, mikrosimülasyon bulgularıyla uyumlu sonuçlar ortaya koymaktadır. Single point kavşak çözümünün, mevcut dönel adalı hemzemin kavşağa kıyasla trafik akımlarını daha düzenli hâle getirdiği, gecikme ve bekleme sürelerini azalttığı ve kapasiteyi artırdığı belirlenmiştir. Özellikle ana arter üzerinde, trafiğin yoğun saatlerde yaşanan trafik sıkışıklığının azaltılmasına önemli katkı sağlamaktadır. Bu yönüyle proje, kavşak tasarım optimizasyonunun sahadaki başarılı bir uygulaması olarak değerlendirilmektedir.



Şekil 13. Uygulama sonrası kavşak alanına ait 01.01.2026 – 30.04.2026 tarihleri arasındaki kaza yoğunluk haritası

Şekil 13'te sunulan uygulama sonrası veriler, kavşak ve yakın çevresindeki trafik kazalarının mekânsal dağılımını ve özet istatistiklerini ortaya koymaktadır. İncelenen dört aylık dönemde toplam 10 kaza meydana gelmiş olup, 5'inin yaralanmalı ve 5'ünün maddi hasarlı olduğu belirlenmiştir. Bu durum, proje öncesine kıyasla kaza sayısı ve şiddetinde önemli bir azalma olduğunu göstermekte ve uygulanan kavşak düzenlemesinin trafik güvenliği üzerinde olumlu bir etki oluşturduğunu ortaya koymaktadır.

Mekânsal analizler, kazaların kavşak merkezinden ziyade bağlantı ve geçiş bölgelerinde yoğunlaştığını göstermekte ve kavşak içindeki çatışma noktalarının azaldığına işaret etmektedir. Ayrıca kazaların sınırlı sayıda ve düşük şiddette gerçekleşmesi, yeni geometrik düzen ve sinyalize işletim yapısının güvenlik performansını iyileştirdiğini desteklemektedir. Bununla birlikte, yaralanmalı kazaların bir kısmının sürücü davranışlarına bağlı olarak ortaya çıkması, yeni kavşak düzenine adaptasyon sürecinin de güvenlik üzerinde etkili olabileceğini düşündürmektedir.

TABLO 3. Single Point Urban Interchange (SPUI) ve Mevcut Hemzemin Kavşak Tasarımlarının Karşılaştırmalı Performans Sonuçları

Performans parametreleri	*Eski tasarım*	*Uygulanan tasarım (SPUI)*
Hizmet düzeyi (LOS)	LOS F	LOS E
Toplam taşıt sayısı (1 saat)	3.804	5.307
Ortalama bekleme süresi (saniye)	158	79
Ortalama hız (km/sa)	6	11

PTV VISSIM mikrosimülasyon analizleri sonucunda, önerilen SPUI tasarımının mevcut dönel adalı kavşak düzenine kıyasla tüm temel performans göstergelerinde belirgin iyileşmeler sağladığı belirlenmiştir. Elde edilen bulgular, kavşak kapasitesinin artırıldığını, trafik akışının daha düzenli ve güvenli hale geldiğini ve taşıt işletim performansının önemli ölçüde iyileştiğini göstermektedir.

Analiz sonuçlarına göre, mevcut durumda LOS F düzeyinde hizmet veren kavşağın, SPUI tasarımı sonrasında LOS E düzeyine yükseldiği görülmüş ve kavşağın işletim koşullarında önemli bir iyileşme sağlandığını ortaya koymaktadır. Bir saatlik analiz periyodunda kavşaktan geçen toplam taşıt sayısının 3.804'ten 5.307'ye yükselmesi, yeni tasarımın daha yüksek trafik hacimlerini karşılayabildiğini göstermekte, ortalama bekleme süresinin 158 saniyeden 79 saniyeye düşmesi ve ortalama hızın 6 km/sa'den 11 km/sa'ye yükselmesi ise kavşağın işletim verimliliğinde belirgin bir iyileşme olduğunu göstermektedir.

Ekonomik ve çevresel analiz

Simülasyon modelinden elde edilen yakıt tüketimi ve gecikme verilerine dayalı olarak yapılan fayda-maliyet analizi, önerilen SPUI tasarımının ekonomik ve çevresel açıdan önemli kazanımlar sağladığını ortaya koymaktadır. Analiz sonuçları aşağıdaki tablo 4'te yer almaktadır. Buna göre, yakıt tüketiminde %29 oranında tasarruf sağlandığı belirlenmiştir. Bu tasarrufun saatlik ekonomik karşılığı 692.523,6 USD, günlük ekonomik karşılığı ise 4.156.380,00 USD olarak hesaplanmıştır. Yıllık bazda değerlendirildiğinde toplam tasarrufun 997.531.200,00 USD düzeyine ulaştığı öngörülmektedir. Ayrıca, çevresel sürdürülebilirlik açısından günlük 1.611 litre yakıt tasarrufu sağlandığı; bunun da günlük yaklaşık 33,10 kg emisyon azalmasına ve dolaylı olarak günde 3,4 ağacın yeniden kazanımına eşdeğer bir çevresel fayda oluşturduğu belirlenmiştir. Bu bulgular, kentsel kavşak iyileştirmelerinin yakıt tüketimi, ekonomik verimlilik ve çevresel etkiler bakımından önemli avantajlar sağlayabildiğini göstermektedir.

TABLO 4. Single Point Urban Interchange (SPUI) Tasarımına İlişkin Ekonomik ve Çevresel Analiz Sonuçları

Gösterge	*Değer*
Yakıt tasarrufu oranı	%29
Saatlik ekonomik fayda	692.523,6 USD
Günlük ekonomik fayda	4.156.380,00 USD
Yıllık ekonomik fayda	997.531.200,00 USD
Günlük yakıt tasarrufu	1.611 litre
Günlük emisyon azalımı	33,10 kg CO_2
Eşdeğer çevresel fayda	3,4 ağaç/gün

SPUI tasarımı ve simülasyon sonuçları

SPUI projesinin 2026 Ocak ayında hizmete alınmasının ardından üç aylık kaza verileri incelendiğinde, ölümlü kazanın meydana gelmediği, 5 yaralanmalı ve 5 maddi hasarlı kazanın gerçekleştiği belirlenmiştir. Bulgular, SPUI tasarımının çatışma noktalarını azaltarak ve trafik akışını daha kontrollü hâle getirerek güvenliği artırdığını göstermekte; özellikle ölümlü kazaların ortadan kalkması bu iyileşmeyi desteklemektedir. Genel olarak sonuçlar, SPUI tasarımının hem işletim hem de trafik güvenliği açısından etkili bir çözüm sunduğunu ortaya koymaktadır.

Bu çalışma, Adana'da birbirinden farklı iki kavşak sorununa mikrosimülasyon destekli çözümler geliştirilmesini

belgelemektedir. PTV VISSIM yazılımının kullanıldığı birinci vakada, hemzemin bir kavşağın SPUI tasarımına dönüştürülmesi somut ve ölçülebilir iyileşmeler sağlamıştır: hizmet düzeyi LOS_F'den LOS_D'ye yükselmiş, saatlik kapasite yaklaşık %72 artmış, ortalama bekleme süresi 166'dan 54 saniyeye gerilemiş, ortalama hız 10'dan 25 km/saate çıkmış ve yakıt tasarrufu %44 olarak gerçekleşmiştir. İkinci vakada ise dönel adalı kavşak üzerinde gerçekleştirilen fiziksel düzenleme; mevcut tıkanıklık koşullarını gidererek kaza riskini azaltmıştır.

Bu bulgular şu temel sonuçlara işaret etmektedir:

- Mikrosimülasyon modellemesi, kavşak tasarım kararlarında güvenilir ve kanıta dayalı bir temel oluşturmaktadır.
- SPUI tasarımı, yüksek trafikli kentsel kavşaklarda hem kapasite hem de güvenlik açısından üstün performans sergilemektedir.
- Kavşak optimizasyonu yakıt tüketiminde ve emisyonlarda kayda değer çevresel kazanımlar sağlamaktadır.
- Veri odaklı mühendislik yaklaşımları, yerel yönetimler açısından somut ve ölçülebilir faydalar üretmektedir.

Gelecek araştırmalar için güvenlik etkilerinin değerlendirilebilmesi amacıyla daha uzun süreli kaza takibinin sürdürülmesi, yaşayan laboratuvar yaklaşımıyla elde edilen simülasyon verilerinin sürekli gerçek zamanlı verilerle güncellenmesi ve bu yöntemlerin Adana'daki diğer sorunlu kavşaklara yaygınlaştırılması önerilmektedir. Bu çalışma, gelişmekte olan Türk kentlerinde kavşak tasarımı ve kentsel trafik yönetimine ilişkin uygulama temelli bir kaynak niteliği taşımaktadır.

TEŞEKKÜR

Bu çalışma, Adana Büyükşehir Belediyesi'nin destek ve katkılarıyla gerçekleştirilmiştir. Veri temini ve saha çalışmalarındaki yardımlarından dolayı Adana Büyükşehir Belediyesi yetkililerine teşekkür ederiz.

KAYNAKLAR

[1] D. Schrank and T. Lomax, 2004 Urban Mobility Report, Texas Transportation Institute, Texas A&M University System, 2004.
[2] R. Elvik, A. Høye, T. Vaa, and M. Sørensen, The Handbook of Road Safety Measures, 2nd ed. Emerald Group Publishing, 2009.
[3] National Highway Traffic Safety Administration, "Traffic safety facts 2018: A compilation of motor vehicle crash data," U.S. Department of Transportation, 2020.
[4] M. Fellendorf and P. Vortisch, "Microscopic traffic flow simulator VISSIM," in Fundamentals of Traffic Simulation, J. Barceló, Ed. Springer, 2010, pp. 63–93. doi: 10.1007/978-1-4419-6142-6_2.
[5] G. Dalkic, M. Balaban, H. Tuydes-Yaman, and T. Celikkol-Kocak, "Design and evaluation of a new intersection model to minimize congestions using VISSIM software," Open Engineering, vol. 10, no. 1, pp. 104–115, 2020. doi: 10.1515/eng-2020-0019.
[6] T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı Coğrafi Bilgi Sistemleri Genel Müdürlüğü, "Trafik Güvenliği Analiz Platformu," 2023. [Online]. Available: https://cbs.csb.gov.tr/trafik-guvenligi-analiz-platformu
[7] T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, "Trafik Güvenliği Analiz Platformu (TGAP)," 2023. [Online]. Available: https://akillisehirler.csb.gov.tr/tgap/
[8] Adana Büyükşehir Belediyesi, "İmar Planlama Şube Müdürlüğü," 2023. [Online]. Available: https://www.adana.bel.tr/tr/birim-detay/89
[9] Transportation Research Board, Highway Capacity Manual 2010, National Research Council, 2010.

Türk Devletlerinde Yapay Zeka: Toplumsal Algı ve Dijital Egemenlik

Prof. Dr. Ebulfez Süleymanlı
Türk Akademisi, Astana, Kazakistan/ Üsküdar Üniversitesi, Sosyoloji Bölümü, İstanbul, Türkiye

e-posta: ebulfez.suleymanli@uskudar.edu.tr

https://orcid.org/ 0000-0003-1894-5232

Doç. Dr. Barış Çağırkan
İzmir Demokrasi Üniversitesi, Sosyoloji Bölümü
İzmir, Türkiye

e-posta: baris.cagirkan@idu.edu.tr

https://orcid.org/0000-0002-0013-1831

Artificial Intelligence in the Turkic States : Social Perception and Digital Sovereignty

Abstract

This study examined the preliminary findings of the Turkic Barometer 2026, a research project conducted by the Turkic Academy that focuses on artificial intelligence-based digital transformation across the Turkic states. Covering Azerbaijan, Kazakhstan, Kyrgyzstan, Uzbekistan, and Türkiye, the study adopted a mixed-methods research design that combined quantitative survey data with semi-structured expert interviews. The preliminary findings indicated that artificial intelligence has become a widely used everyday tool among different social and occupational groups. The results also revealed considerable uncertainty regarding employment risks, limited public awareness of national artificial intelligence investments, and differentiated patterns of digital adaptation across the five countries. While Türkiye and Kazakhstan appeared as relatively proactive actors in terms of public discourse and institutional capacity, Azerbaijan and Kyrgyzstan demonstrated more cautious and capacity-sensitive profiles, and Uzbekistan reflected a transitional model of rapid adoption. The study argues that the Turkic world needs a common artificial intelligence vision grounded in digital sovereignty, ethical governance, regional data cooperation, digital literacy, and Turkic-language large language models. In this context, artificial intelligence constitutes a sociological, cultural, strategic, and regional cooperation challenge as much as a technological one.

Keywords – Artificial intelligence, digital transformation, Turkic world, Turkic Barometer, digital sovereignty.

I. Giriş

Yapay zeka, yirmi birinci yüzyılın en etkili teknolojik dönüşüm alanlarından biri haline gelmiştir [1]. Başlangıçta daha çok teknik uzmanlık, yazılım geliştirme ve mühendislik alanlarıyla ilişkilendirilen yapay zeka, günümüzde gündelik hayatı, çalışma kültürünü, eğitim sistemlerini, kamu yönetimini, güvenlik politikalarını, medya ekosistemini ve devletlerin stratejik gelecek tasavvurlarını doğrudan etkilemektedir. Bu nedenle yapay zeka, teknik bir yenilik olmanın ötesinde, toplumsal, kültürel, ekonomik ve siyasal sonuçları olan çok boyutlu bir dönüşüm alanıdır. Sosyolojik açıdan bakıldığında yapay zeka, algoritmalardan ve dijital araçlardan oluşan bir teknoloji setinden ibaret değildir; yapay zeka, bireylerin bilgiye erişme biçimini, kurumların karar alma süreçlerini, emek piyasasının yapısını [15], mahremiyet algısını, güven ilişkilerini ve toplumsal denetim mekanizmalarını dönüştüren yeni bir sosyo-teknik gerçekliktir [16]. Bu nedenle

yapay zeka tartışmaları, teknolojik kapasiteyle birlikte etik, hukuk, eğitim, kültür, istihdam, veri egemenliği ve bölgesel iş birliği boyutlarıyla ele alınmalıdır.

Yapay zekanın yarattığı bu sosyo-teknik gerçeklik, coğrafi ve kültürel bağlamlara göre farklı biçimlerde tezahür etmekte; bölgesel düzeyde özgün dinamikler üretmektedir. Türk dünyası açısından yapay zeka temelli dijital dönüşüm bu çerçevede ayrıca stratejik bir anlam taşımaktadır. Türk devletleri ortak tarihî, kültürel ve dilsel bağlara sahip olmasına karşın dijital altyapı, insan kaynağı, kurumsal kapasite ve toplumsal farkındalık bakımından farklı düzeylerde bulunmaktadır. Yapay zeka çağında dışarıdan teknoloji kullanan toplumlar, küresel teknoloji tekellerine ve dış veri altyapılarına daha bağımlı hale gelebilmektedir [12]. Buna karşılık kendi veri ekosistemini kurabilen, dil teknolojilerini geliştirebilen, insan kaynağını dönüştürebilen ve etik yönetişim ilkelerini oluşturabilen ülkeler, dijital egemenlik ve stratejik özerklik bakımından daha güçlü bir konuma ulaşacaklardır.

Yapay zeka ve dijital dönüşüm konularına ilişkin küresel ve ülke düzeyinde çalışmalar artmakla birlikte; Türk dünyasını ortak bir karşılaştırmalı çerçevede ele alan ampirik araştırmaların sayısı sınırlı kalmaktadır. Bu çalışma, Türk Akademisinin yürüttüğü "*Türk Barometresi 2026: Türk Dünyasında Yapay Zeka Temelli Dijital Dönüşüm*" başlıklı projenin ilk bulgularından hareketle, söz konusu dönüşümün toplumsal algılar, kullanım pratikleri, mesleki kaygılar, stratejik beklentiler ve ülke profilleri açısından nasıl şekillendiğini değerlendirmeyi amaçlamaktadır. Çalışma; beş Türk devletini ortak bir karşılaştırmalı çerçevede ele alan ilk barometre temelli ampirik çözümlemeyi sunması ve bulguları kültürel gecikme ile dijital egemenlik literatürleriyle ilişkilendirmesi bakımından özgün bir katkı sunmaktadır. Bu doğrultuda araştırma; Azerbaycan, Kazakistan, Kırgızistan, Özbekistan ve Türkiye'yi kapsayan karşılaştırmalı bir ampirik çerçeve sunmayı ve elde edilen bulgular ışığında Türk dünyası için politika önerileri geliştirmeyi hedeflemektedir.

II. Literatür ve Politika Arka Planı

Yapay zeka yönetişimi, son yıllarda uluslararası kuruluşların ve ulusal hükümetlerin temel gündem maddelerinden birine dönüşmüştür. Bu alanda öncü çerçevelerden birini oluşturan İktisadi İşbirliği ve Kalkınma Örgütü'nün (OECD) 2019 tarihli Yapay Zeka İlkeleri, güvenilir yapay zekanın geliştirilmesi ve yaygınlaştırılması sürecinde insan hakları, demokratik değerler, şeffaflık, sağlamlık, hesap verebilirlik ve kapsayıcı büyüme gibi temel ilkeleri ön plana çıkarmaktadır [2]. Söz konusu ilkeler, üye devletlerin ulusal yapay zeka stratejilerini biçimlendirmede referans çerçevesi işlevi görmekte ve politika yapıcılara somut yönelim sağlamaktadır. Birleşmiş Milletler Eğitim, Bilim ve Kültür Örgütü'nün (UNESCO) Yapay Zeka Etiği Tavsiye Kararı ise yapay zekanın insan onuru, insan hakları, adalet, sürdürülebilirlik, açıklanabilirlik ve toplumsal sorumluluk ilkeleriyle birlikte değerlendirilmesi gerektiğini vurgulamaktadır [3].

Küresel işgücü piyasası tartışmaları da yapay zekanın istihdam, mesleki beceriler ve çalışma kültürü üzerindeki etkisini giderek daha görünür kılmaktadır. Otomasyon ve makine öğrenmesi uygulamalarının yaygınlaşmasıyla birlikte pek çok meslek grubu, görev tanımlarının yeniden şekillendiği bir dönüşüm süreciyle karşı karşıya kalmaktadır. Dünya Ekonomik Forumu'nun 2025 tarihli Future of Jobs raporu; teknolojik değişim, yapay zeka kullanımı, yeşil dönüşüm, demografik değişimler ve jeoekonomik kırılmaların 2025–2030 döneminde işgücü piyasasını şekillendirecek başlıca unsurlar arasında yer aldığını belirtmektedir [4]. Rapora göre analitik düşünme, yaratıcılık, teknolojik okuryazarlık ve esneklik gibi yetkinlikler önümüzdeki dönemde en çok talep görecek beceriler arasında öne çıkarken; rutin ve tekrara dayalı görevlerin otomasyona devredilme hızı artmaktadır. Bu tablo, mesleki becerilerin statik değil dinamik bir yapı kazandığını ve işgücünün sürekli olarak kendini yenilemesi gerektiğini ortaya koymaktadır. Dolayısıyla yaşam boyu öğrenme, yeniden beceri kazanımı ve dijital yetkinlikler; bireysel kariyer stratejilerinin ötesinde ulusal rekabet gücünü ve toplumsal dayanıklılığı doğrudan etkileyen stratejik politika başlıkları haline gelmektedir.

Türk devletlerinde de yapay zeka ve dijital dönüşüm alanında politika belgeleri ve stratejik yönelimler gelişmektedir. Türkiye'nin Ulusal Yapay Zeka Stratejisi 2021–2025 belgesi; insan kaynağı, araştırma ve inovasyon, kaliteli veriye erişim, teknik altyapı, sosyoekonomik uyum ve uluslararası iş birliği gibi başlıkları öncelemektedir [5]. Kazakistan'ın 2024–2029 Yapay Zeka Geliştirme Konsepti; veri altyapısı, süper bilgisayar kapasitesi, ulusal yapay zeka platformu ve kamu süreçlerine entegrasyon gibi alanlara odaklanmaktadır [6]. Özbekistan'ın 2030'a kadar yapay zeka teknolojilerinin geliştirilmesine yönelik stratejisi, yapay zekanın ekonomik ve sosyal sektörlerde yaygınlaştırılmasını hedeflemektedir [7]. Azerbaycan'ın 2025–2028 Yapay Zeka Stratejisi ve Kırgızistan'ın 2024–2028 Dijital Dönüşüm Konsepti de bölgesel ölçekte yapay zeka ve dijitalleşme gündeminin güçlendiğini göstermektedir [8], [9]. Politika belgelerinin varlığı, toplumsal hazırlığın kendiliğinden gerçekleştiği anlamına gelmemektedir. Asıl mesele, bu stratejilerin toplum tarafından ne ölçüde bilindiği, benimsendiği ve gündelik pratiklere nasıl yansıdığıdır. Türk Barometresi 2026'nın ilk bulguları, tam da bu politika vizyonu ile toplumsal uyum arasındaki ilişkiyi anlamak açısından önem taşımaktadır.

III. Yöntem

A. Araştırmanın Deseni

Bu çalışma, Türk Akademisi tarafından yürütülen Türk Barometresi 2026 projesinin ilk bulgularına dayanmaktadır. Araştırma; nicel ve nitel verilerin birlikte kullanıldığı karma yöntem deseniyle tasarlanmıştır. Karma yöntem yaklaşımı; toplumsal algıların sayısal dağılımını ölçme ile bu algıların altında yatan anlam örüntülerini derinlemesine çözümleme işlevlerini tek bir araştırma çerçevesinde bir araya getirdiği için tercih edilmiştir. Bu çerçevede nicel boyut, beş ülke arasındaki

karşılaştırmalı betimlemeyi olanaklı kılmış; nitel boyut ise ülke profillerinin ayrıntılı yorumlanmasına ve sayısal örüntülerin bağlamsal açıklamasına olanak sağlamıştır. İki yöntemsel boyut, araştırma sorularını ortak bir kavramsal çerçevede ele alacak biçimde bütünleştirilmiştir.

B. Evren, Örneklem ve Veri Toplama Araçları

Araştırmanın nicel boyutunda Azerbaycan, Kazakistan, Kırgızistan, Özbekistan ve Türkiye'nin her birinden yaklaşık 1.070 katılımcıya ulaşılması hedeflenmiş; beş ülke genelinde 5.000'in üzerinde anket verisinin toplanması planlanmıştır. Katılımcılar farklı meslek gruplarından seçilmiştir. Anket formu; yapay zeka kullanım sıklığı, yapay zekaya yönelik duygular, mesleki gelecek ve işsizlik kaygısı, ulusal yatırımlara ilişkin algı ve dijital strateji tercihleri başlıkları üzerinden yapılandırılmıştır. Araştırmanın nitel boyutunda; yapay zeka, siber güvenlik, eğitim teknolojileri, dijital dönüşüm ve teknoloji politikaları alanlarında çalışan uzmanlarla yarı yapılandırılmış derinlemesine görüşmeler gerçekleştirilmiştir. Her ülkeden 30 uzman olmak üzere toplam 150 uzmanla görüşme yapılmıştır. Görüşmeler; ülkelerin dijital uyum profillerini, insan kaynağı kapasitesini, siber güvenlik algılarını, veri egemenliği kaygılarını ve bölgesel iş birliği olanaklarını anlamayı amaçlamıştır.

C. Verilerin Analizi

Araştırmada elde edilen veriler, karma yöntem desenine uygun biçimde iki aşamalı bir analiz sürecinden geçirilmiştir. İlk aşamada nicel ve nitel veriler kendi içinde bağımsız olarak çözümlenmiş; ikinci aşamada bu iki veri kümesinden elde edilen bulgular karşılaştırmalı ve bütünleşik bir çerçevede yorumlanmıştır. Böylece her iki veri kümesinin analitik özerkliği korunmuş, bütünleştirme aşamasında ise örtüşen ve ayrışan örüntüler daha güçlü bir görgül zemin üzerinde karşılaştırılmıştır. Bu yaklaşım, sayısal eğilimlerin ardındaki anlam örüntülerini görünür kılmayı, ülke profillerini çok katmanlı biçimde değerlendirmeyi ve tek bir veri kaynağına dayalı çıkarımların taşıdığı yorum sınırlılıklarını azaltmıştır.

Nicel veriler betimsel istatistik teknikleriyle çözümlenmiştir. Katılımcı yanıtları; frekans, yüzde ve ortalama dağılımları üzerinden ülke, meslek grubu, yaş ve cinsiyet değişkenleri bazında karşılaştırılmıştır. Nitel veriler tematik içerik analizi yoluyla değerlendirilmiştir. Görüşme kayıtları öncelikle deşifre edilmiş ve anlam birimlerine ayrılmıştır. Bu anlam birimleri açık kodlama tekniğiyle etiketlenmiş, kodlardan alt temalar ve ana temalar oluşturulmuştur. Tematik çerçeve; dijital uyum, insan kaynağı kapasitesi, siber güvenlik algısı, veri egemenliği kaygısı ve bölgesel iş birliği olanakları olmak üzere beş ana eksen üzerinde yapılandırılmıştır. Son aşamada nicel ve nitel bulgular üçgenleme yaklaşımıyla karşılaştırılmıştır. Sayısal eğilimlerin uzman görüşleriyle hangi noktalarda örtüştüğü, hangi noktalarda ayrıştığı belirlenmiş; böylece birbirini doğrulayan ve tamamlayan bütünleşik bir yorum çerçevesi oluşturulmuştur. Bu çerçeve, ülke profillerinin karşılaştırmalı çözümlemesine ve Türk dünyası ölçeğinde genel örüntülerin ortaya konulmasına temel oluşturmuştur.

D. Etik Hususlar ve Sınırlılıklar

Araştırma, insan katılımcılarla yürütülen sosyal bilim çalışmalarına ilişkin etik standartlara uygun biçimde tasarlanmıştır. Bu kapsamda çalışma için İzmir Demokrasi Üniversitesi Sosyal ve beşeri Bilimler Bilimsel Araştırma ve Yayın Etik Kurulundan 2026/04-08 karar sayı ile etik onay alınmıştır. Veri toplama sürecinde gönüllü katılım ilkesi esas alınmış; tüm katılımcılara araştırmanın amacı, kapsamı, veri kullanım biçimi ve haklarına ilişkin bilgi verilerek bilgilendirilmiş onamları alınmıştır. Katılımcılar, araştırmadan herhangi bir aşamada çekilme hakkına sahip olduklarına ilişkin açıkça bilgilendirilmiştir.

Verilerin işlenmesinde anonimlik ve gizlilik ilkeleri gözetilmiştir. Anket verileri kimliksizleştirilerek saklanmış; nitel görüşmelerde elde edilen kayıtlarda katılımcıların kimliğini doğrudan ya da dolaylı biçimde açığa çıkarabilecek bilgiler analiz öncesinde ayıklanmıştır. Veriler güvenli ortamlarda saklanmış ve sadece araştırma ekibinin erişimine açık tutulmuştur. Yapay zeka odaklı bir araştırma olması nedeniyle çalışma; veri mahremiyeti, algoritmik şeffaflık ve sorumlu araştırma uygulamaları gibi konularda ek bir duyarlılıkla yürütülmüştür.

Bu makalede sunulan veriler, Türk Barometresi 2026 araştırmasının ilk bulgularına dayanmaktadır. Veri toplama süreci sürdüğünden, sonuçlar ön bulgular olarak değerlendirilmelidir; nihai sonuçlar olarak okunmamalıdır. Bulguların beş ülkeyle sınırlı olması ve örneklemin meslek gruplarına göre dağılımı, sonuçların genellenebilirliği açısından dikkate alınmalıdır. Ayrıca öz-bildirime dayalı anket verilerinin sosyal beğenirlik eğilimi taşıyabileceği, ülkeler arasındaki dijital altyapı ve kültürel bağlam farklarının karşılaştırmalı yorumlarda göz önünde bulundurulmalıdır.

IV. İlk Bulgular

A. Yapay Zeka Kullanım Sıklığı

Araştırmanın ilk bulguları, yapay zeka kullanımının Türk dünyasında farklı toplumsal ve mesleki gruplara yayılmış yüksek bir yoğunluk sergilediğini ortaya koymuştur. Katılımcıların %36,2'si yapay zekayı her gün, %30,9'u haftada birkaç kez kullandığını belirtmiştir. Buna karşılık yapay zekayı hiç kullanmadığını ifade edenlerin oranı %4,4 düzeyinde kalmıştır. Bu iki kategorinin toplamı, katılımcıların yaklaşık %67,1'inin yapay zekayı düzenli kullanıcı olarak konumlandığını; kullanmayanların ise oldukça küçük bir azınlık oluşturduğunu göstermiştir. Bu dağılım, yapay zekanın kullanıcı tabanının sınırlı bir uzman grubunun ötesine geçtiğine ve genel nüfusa yayılmış bir kullanım biçimi kazandığına işaret etmektedir.

Elde edilen kullanım yoğunluğu, yapay zekanın uzmanlık temelli teknik bir araç olmaktan çıkıp gündelik hayatın sıradan bir bileşeni haline geldiğine işaret etmektedir. Sosyolojik açıdan bu göstergeler; yapay zekanın "yenilikçi ürün" aşamasından "kültürel olarak normalleşmiş teknoloji"

aşamasına geçiş sürecini yaşadığını göstermektedir. Öte yandan kullanım sıklığının yüksek olması, kullanımın niteliğine ilişkin doğrudan bir gösterge oluşturmamaktadır. Yoğun kullanım; bilinçli, güvenli ve etik bir kullanım anlamına gelebileceği gibi yüzeysel, sorgulanmadan benimsenmiş ve riskleri görünmez kılan bir kullanıma da karşılık gelebilmektedir.

Bu çerçevede yaygın kullanım örüntüsünün, yapısal düzeyde çeşitli risk alanlarını beraberinde getirdiği görülmektedir. Öne çıkan başlıklar; veri mahremiyetinin aşınması, kaynak doğrulama pratiklerinin zayıflaması, dezenformasyona karşı kırılganlık, algoritmik yanlılığın bireysel kararlar üzerindeki görünmez etkisi ve dış platformlara ya da yabancı model sağlayıcılarına oluşan yapısal bağımlılıktır. Bu risklerin ortak özelliği, kullanım sıklığı arttıkça görünürlüğünün azalmasıdır. Bu örüntü, araştırmanın tartışma bölümünde de ele alınan kültürel gecikme dinamiğini somutlaştırmaktadır: Yapay zeka araçlarının teknik yayılımı, bu araçları sorumlu biçimde kullanacak kurumsal kapasite, hukuki düzenleme ve etik farkındalıkla aynı hızda ilerlememektedir. Dolayısıyla Türk dünyasında yapay zekanın yaygın kullanımı; başlı başına olumlu bir kapasite göstergesi olarak okunmamalı, dijital okuryazarlık ve etik farkındalık programlarıyla desteklenmesi gereken bir durum olarak değerlendirilmelidir.

B. Yapay Zekaya Yönelik Duygular

Araştırmanın duygu haritası, Türk dünyasında yapay zekaya yönelik baskın duygunun korku ya da reddediş değil, keşif odaklı bir merak olduğunu ortaya koymuştur. Katılımcıların %65'i yapay zeka karşısında merak, %42'si heyecan duyduğunu belirtmiştir. Olumsuz duygular daha düşük düzeylerde kalmış; belirsizlik %28, endişe %22 ve korku %12 oranında ölçülmüştür. Bu dağılım, olumlu duyguların olumsuz duyguların yaklaşık iki katı düzeyinde seyrettiğini; toplumsal duygu profilinin açık biçimde keşif ekseninde yoğunlaştığını göstermiştir.

Elde edilen duygu profili, Türk dünyasında yapay zekaya yönelik savunmacı ve kaygı merkezli bir atmosferin baskın olmadığını göstermiştir. Sosyolojik açıdan bu örüntü; yeni bir teknolojiye yönelik toplumsal karşılaşma biçiminin direnci yerine öğrenme, deneme ve uyum sağlama eğilimi üzerinde şekillendiğini düşündürmektedir. Keşif odaklı merakın baskın duygu olarak öne çıkması, kullanıcıların yapay zekayı dışsal ve tehditkar bir araç yerine bireysel kapasite genişlemesi olarak algıladıklarına işaret etmektedir. Bu durum, kamu politikaları açısından önemli bir fırsat alanı oluşturmaktadır; merak duygusunun baskın olduğu bir dönemde yapılandırılacak dijital okuryazarlık ve etik farkındalık programları, yapay zekanın daha bilinçli, güvenli ve üretken biçimde benimsenmesini destekleyebilir.

Bununla birlikte baskın meraka indirgenemeyen ikincil bir duygu kümesinin varlığı görünmektedir. Belirsizlik, endişe ve korku oranları birlikte değerlendirildiğinde, katılımcıların yaklaşık %62'sinin yapay zeka karşısında en az bir olumsuz duygu beyan ettiği görülmektedir. Bu kesim, teknolojiye toptan karşı olan bir niş grup oluşturmamakta; merakla birlikte yaşanan ikircikli

(ambivalan) bir duygu durumunu yansıtmaktadır. Sosyolojik literatürde bu tür bir duygu yapısı, yeni teknolojilerin sunduğu olanaklar ile yarattığı belirsizlikler arasında kalan toplumlara özgü tipik bir uyum aşaması olarak değerlendirilmektedir. Duygu haritasındaki bu ikili yapı, yapay zeka politikalarının teknik yatırım ve altyapı boyutuyla sınırlı kalmaması gerektiğini göstermektedir. Yapay zekanın toplumsal etkileri şeffaf, anlaşılır ve erişilebilir bir dille paylaşılmadığı takdirde, merak duygusunun zamanla güvensizliğe, kuşkuya ve geri çekilmeye dönüşme riski bulunmaktadır. Bu nedenle yapay zeka yönetişiminin toplumsal iletişim boyutu; risk iletişimi, kamuoyu bilgilendirmesi ve toplumsal katılım mekanizmalarıyla birlikte teknik yatırım boyutu kadar stratejik bir önem taşımaktadır.

C. Mesleki Gelecek ve İşsizlik Kaygısı

Yapay zeka tartışmalarının en somut etki alanlarından biri çalışma hayatıdır. Araştırmanın bulguları, Türk dünyasında yapay zekanın istihdam üzerindeki etkisine ilişkin belirgin bir kararsızlık bulunduğunu ortaya koymuştur. Katılımcıların %31'i yapay zekanın işsizlik riski konusunda kararsız kalmıştır. Bu oran, beyan düzeyinde açık bir kanaate ulaşamamış en geniş tekil grubun kararsızlar olduğuna işaret etmektedir. Toplumun önemli bir kesimi; yapay zekanın kendi mesleğini ortadan kaldıracağı, dönüştüreceği ya da kolaylaştıracağı yönünde net bir konum geliştirememiştir. Bu dağılım, yapay zekanın istihdam etkilerine ilişkin belirsizliğin bireysel bir tereddütten çok toplumsal düzeyde paylaşılan bir epistemik boşluk olarak deneyimlendiğini göstermiştir.

Elde edilen kararsızlık örüntüsü, yapay zekanın çalışma hayatı üzerindeki etkisinin toplumsal düzeyde henüz net bir anlatıya kavuşturulamadığını göstermektedir. Sosyolojik açıdan bu durum, bireylerin yapay zekanın meslekleri ne yönde dönüştüreceğini değerlendirebileceği güvenilir bir referans çerçevesinden yoksun kalmasıyla açıklanabilir. Mesleki gelecek algısı, kişisel beklentilerin yanı sıra kamu kurumlarının, meslek örgütlerinin, eğitim sisteminin ve medyanın ürettiği anlatılardan beslenmektedir. Bu anlatıların henüz yeterince oluşmamış olması, bireylerin geleceğe ilişkin senaryo üretme kapasitesini sınırlamakta ve kararsızlığı yapısal bir eğilim haline getirmektedir.

Kararsızlık, aynı zamanda emek piyasasının dönüşüm dinamiğiyle de yakından ilişkilidir. Yapay zeka; bazı görevleri otomatikleştirerek ortadan kaldırmakta, bazı meslekleri yeniden tanımlamakta ve yeni iş alanları yaratmaktadır. Bu üç eğilim aynı anda işlediğinden, bireylerin kendi meslekleri için net bir gelecek tahmini yapması güçleşmektedir. Kararsızlık bu nedenle bilgi eksikliğiyle birlikte dönüşümün çok yönlü ve eş zamanlı karakterinin bir yansıması olarak okunmalıdır. Bu durum, mesleki rehberlik, yeniden beceri kazanımı ve beceri yükseltme politikalarının stratejik önemini artırmaktadır.

Yapay zeka çağında asıl soru, hangi mesleklerin kaybolacağından çok mevcut mesleklerin hangi yeni becerilerle yeniden tanımlanacağıdır. Dolayısıyla Türk dünyasında yapılacak öncelikli yatırım; mesleki dönüşüm haritalarının hazırlanması, ortak dijital beceri geliştirme programlarının oluşturulması, çevrim içi eğitim platformlarının yaygınlaştırılması ve yapay zeka destekli kariyer rehberliği modellerinin geliştirilmesi olmalıdır. Üniversiteler, kamu kurumları, meslek örgütleri ve özel sektör arasında kurulacak yapısal bir iş birliği, mevcut kararsızlığı dönüşe hazır bir kapasiteye çevirmenin temel koşulu olarak öne çıkmaktadır.

D. Stratejik Önem ve Yatırım Algısı

Araştırmanın bulguları, Türk dünyasında yapay zekanın stratejik anlamına ve yatırım düzeyine ilişkin algıların belirgin biçimde parçalı bir görünüm sergilediğini ortaya koymuştur. Katılımcıların %41,6'sı yapay zekanın küresel rekabet gücü açısından stratejik olduğunu düşündüğünü belirtmiş; %36,9'u bu konuda kararsız kalmıştır. Benzer biçimde "ülkem yapay zekaya yeterli yatırım yapıyor" ifadesine katılımcıların %28,7'si olumlu yanıt verirken, %42,3'ü kararsız kalmış, %29'u yatırımları yetersiz bulmuştur. İki sorunun ortak örüntüsü, kararsız grubun her iki boyutta da en geniş tekil grup konumuna yaklaşması; net stratejik ya da net yatırım kanaatine ulaşan kesimlerin ise azınlıkta kalmasıdır.

Elde edilen bu dağılım, yapay zekanın gündelik kullanım düzeyinde yaygınlaşmış olmasına karşın stratejik anlamının toplum tarafından henüz yeterince içselleştirilmediğini göstermektedir. Sosyolojik açıdan bu durum, "araç olarak yapay zeka" ile "stratejik kaynak olarak yapay zeka" arasındaki algısal mesafenin hâlâ açık olduğunu göstermektedir. Bireyler yapay zekayı günlük pratiklerinde sık biçimde kullanmakta; ancak aynı teknolojinin ulusal rekabet gücü, ekonomik egemenlik ya da teknolojik bağımsızlık açısından taşıdığı anlamı değerlendirebileceği bir bilgi çerçevesinden yoksun kalmaktadır. Yatırım algısında öne çıkan yüksek kararsızlık oranı da bunu desteklemekte; ulusal yapay zeka stratejilerinin, kamu yatırım programlarının ve özel sektör girişimlerinin toplum nezdindeki görünürlüğünün sınırlı kaldığını düşündürmektedir.

Bu çerçevede stratejik algı ile gündelik kullanım arasındaki açıklığın yapısal düzeyde çeşitli risk alanlarını beraberinde getirdiği görülmektedir. Öne çıkan başlıklar; ulusal yapay zeka stratejilerinin kamuoyu desteğinden yoksun kalması, yatırım kararlarının toplumsal meşruiyet zemininin zayıflaması, stratejik tartışmaların dar uzman çevreleriyle sınırlı kalması ve bilgi açığının yabancı platformların ürettiği anlatılarla doldurulmasıdır. Bu risklerin ortak özelliği, stratejik farkındalık oluşmadan yapılan yatırımların toplumsal karşılık bulmasının güçleşmesi; dolayısıyla teknik yatırım kapasitesi ile toplumsal kapasite arasında yapısal bir uyumsuzluğun oluşmasıdır. Dolayısıyla yapay zeka politikaları uzman raporları ile devlet belgelerinde sınırlı kalmamalı; kamu kurumları, üniversiteler ve medya yapay zekanın stratejik önemini toplumsal düzeyde anlatacak açık, anlaşılır ve sürekli bir iletişim dili geliştirmelidir.

E. Dijital Strateji Vizyonu

Araştırmanın bulguları, Türk dünyasında katılımcıların tercih ettiği dijital strateji vizyonunun tek bir eksende yoğunlaşmadığını; çok eksenli ve dengeli bir dağılım sergilediğini ortaya koymuştur. Teknoloji öncülüğü ve inovasyon %28,9 ile ilk sırada yer almış; dijital egemenlik ve güvenlik %27,6 ile ikinci sırada gelmiştir. Sosyal fayda ve etik %23, uluslararası iş birliği ise %20,5 oranında tercih edilmiştir. Dört eksen arasındaki farkın dar bir aralıkta (yaklaşık 8,4 puan) kalması; herhangi bir vizyonun diğerlerini dışlayacak biçimde baskın hale gelmediğini, toplumun dijital geleceğe ilişkin beklentilerini birbirini tamamlayan eksenler üzerinden tanımladığını göstermiştir.

Elde edilen bu dağılım, Türk dünyasında yapay zekaya yönelik beklentilerin tek boyutlu bir "ilerleme" ya da "güvenlik" anlatısına indirgenemeyeceğini göstermektedir. Sosyolojik açıdan dört eksenin yakın oranlarda tercih edilmesi; toplumun dijital dönüşümü tek boyutlu bir teknoloji yarışının ötesinde, çok katmanlı bir egemenlik, güvenlik ve sorumluluk meselesi olarak algıladığına işaret etmektedir. Teknoloji öncülüğü ile dijital egemenliğin neredeyse eşit oranda tercih edilmesi özellikle dikkat çekicidir: bu örüntü, katılımcıların inovasyonu müstakil bir hedef olarak değil; veri güvenliği, siber bağımsızlık ve teknolojik egemenlikle birlikte düşünülmesi gereken bir kapasite olarak konumlandırdığını göstermektedir. Sosyal fayda-etik ile uluslararası iş birliği eksenlerinin de önemli birer pay almış olması; toplumun yapay zekanın ulusal kapasiteye yönelik etkileriyle birlikte bölgesel ve küresel etkilerini de takip ettiğine işaret etmektedir.

Bu çerçevede dört eksenin dengeli dağılımının, dijital strateji tasarımı açısından çeşitli yapısal risk ve fırsat alanlarını beraberinde getirdiği görülmektedir. Öne çıkan başlıklar; tek eksene odaklanan stratejilerin toplumsal meşruiyet aralığını daraltma riski, dijital egemenlik vurgusunun yeterince karşılanmaması durumunda oluşabilecek yabancı platformlara bağımlılık kırılganlığı, etik-sosyal fayda boyutunun ihmal edilmesi halinde teknoloji-toplum arasında oluşabilecek güven aşınması ve uluslararası iş birliği boyutunun zayıf kalması halinde Türk dünyasının küresel yapay zeka yönetişimi tartışmalarının dışında kalma olasılığıdır. Bu risklerin ortak özelliği, dört eksenden herhangi birinin ihmal edilmesi durumunda diğer eksenlerde elde edilen kazanımların da aşınmasıdır.

Dolayısıyla Türk dünyası için geliştirilecek yapay zeka vizyonu; inovasyon kapasitesi, dijital egemenlik, etik-toplumsal sorumluluk ve uluslararası iş birliği eksenlerini birbirinin alternatifi olarak değil, birbirini tamamlayan bileşenler olarak içermelidir. Toplumun dengeli tercih örüntüsü, bu dört eksenli bütünleşik bir strateji yaklaşımı için güçlü bir toplumsal meşruiyet zemini oluşturmaktadır; böyle bir vizyonun tasarlanması ve uygulanması Türk dünyasının dijital gelecekteki konumlanışını belirleyecek temel politika tercihi olarak öne çıkmaktadır.

V. Karşılaştırmalı Ülke Profilleri

Nitel bulgular, araştırma kapsamındaki beş ülkenin yapay zeka ve dijital dönüşüm alanında birbirinden farklı uyum profilleri sergilediğini ortaya koymuştur. Ortaya çıkan profiller; ulusal strateji belgelerinin olgunluk düzeyi, kurumsal kapasite, insan kaynağı, dijital okuryazarlık ve dış teknoloji sağlayıcılarına bağımlılık gibi çok sayıda yapısal değişkenin etkileşimi sonucunda biçimlenmektedir. Bu çerçevede Azerbaycan, Kazakistan, Kırgızistan, Özbekistan ve Türkiye örnekleri; Türk dünyasının yekpare bir dijital dönüşüm modeli içinde değerlendirilemeyeceğini, aksine her ülkenin kendi kurumsal birikimi ve toplumsal dinamikleri tarafından şekillendirilen özgün bir uyum stratejisi izlediğini göstermektedir.

Azerbaycan'ın profili, güçlü bir devlet vizyonu ve göreli olarak olgunlaşmış bir e-devlet tecrübesi ile karakterize olmaktadır. Bununla birlikte uzman görüşleri, insan kaynağı ve eğitim altyapısını yapay zeka alanındaki temel darboğazlardan biri olarak tanımlamaktadır. Veri egemenliğinin korunması, dış teknoloji sağlayıcılara temkinli yaklaşım ve bireysel mahremiyetin güvenceye alınması ise Azerbaycan örneğinde öne çıkan önemli politika başlıkları arasında yer almaktadır. Bu görünüm; Azerbaycan için kurumsal-stratejik olgunluk ile insan kaynağı kapasitesi arasında dengelenmesi gereken bir gelişim alanı bulunduğuna işaret etmektedir.

Kazakistan, araştırma kapsamındaki ülkeler arasında daha sistematik ve kurumsal bir dijital dönüşüm profili sergilemiştir. Ülkenin yapay zeka stratejisi; veri altyapısı, süper bilgisayar kapasitesi, ulusal yapay zeka platformu ve kamu süreçlerine entegrasyon başlıkları çevresinde yapılandırılmıştır [6]. Uzman görüşleri, ülkenin pilot uygulamalar evresinden sistematik entegrasyon evresine geçiş aşamasında bulunduğuna işaret etmektedir. Bununla birlikte kuşaklar arası dijital uçurum ile dış bulut altyapılarına bağımlılık, kurumsal olgunlukla birlikte değerlendirildiğinde önemli yapısal kırılganlık alanları olarak öne çıkmaktadır. Bu çerçevede Kazakistan örneği; güçlü bir kamu kapasitesinin tek başına dijital egemenliği güvenceye almaya yetmediğini, kapasiteyle eş zamanlı olarak teknolojik bağımsızlık yatırımlarının da önem taşıdığını göstermektedir.

Kırgızistan, araştırma kapsamındaki ülkeler arasında daha temkinli ve kırılgan bir profil sergilemektedir. Yapay zeka kullanımının artış gösterdiği gözlemlenmekle birlikte; düşük dijital okuryazarlık düzeyi, dezenformasyona karşı kırılganlık, derin sahte (deepfake) içeriklerin yayılım potansiyeli ve yerel dil temelli yapay zeka altyapısının sınırlılığı, önemli risk alanları olarak belirlenmiştir. Bu durum; dijital okuryazarlığın bireysel düzeyde bir teknik beceri olmanın yanı sıra toplumsal dayanıklılık ve bilgi güvenliğinin yapısal bir bileşeni olarak konumlandırılması gerektiğini göstermektedir. Kırgızistan örneği, yapay zeka çağında dijital okuryazarlık politikalarının ihmal edildiği durumlarda teknolojiye erişimin tek başına toplumsal kazanıma dönüşmediğini ortaya koymaktadır.

Özbekistan, hızlı yayılım ve aktif geçiş evresiyle öne çıkan bir profil sergilemektedir. Ülkenin 2030 yılına kadar yapay zeka teknolojilerini geliştirmeye yönelik ulusal stratejisi, yapay zekanın farklı ekonomik ve toplumsal sektörlerde yaygınlaştırılmasını öngörmektedir [7]. Nitel bulgular; yapay zeka araçlarıyla pratik düzeyde proje geliştiren, buna karşılık derin teknik bilgiye her zaman sahip olmayan yeni kullanıcı tiplerinin ortaya çıktığını göstermiştir. Bu örüntü, Özbekistan'da pragmatik ancak dengesiz bir yayılım modelinin oluşmakta olduğuna; uygulama hızının teknik derinlik ve nitelikli işgücü gelişiminin önünde seyrettiğine işaret etmektedir.

Türkiye, yaygın kullanıcı tabanı ve göreli olarak gelişmiş bir eleştirel söylem kapasitesi ile öne çıkmıştır. Uzman görüşlerinde Türkiye'nin durumu "kullanan ama üretemeyen" paradoksu çerçevesinde değerlendirilmiş; yapay zeka uygulamalarına yönelik toplumsal ilginin güçlü olduğu, buna karşılık bu ilginin yerli üretim, temel model geliştirme ve derin teknoloji kapasitesiyle aynı hızda ilerlemediği vurgulanmıştır. Söz konusu paradoks, Türkiye'nin yapay zeka ekosisteminde tüketim odaklı olgunluğun üretim odaklı olgunluğun ilerisinde seyrettiğine işaret etmekte; ülkenin orta vadeli stratejik konumlanışı açısından yapısal bir kırılganlık alanı oluşturmaktadır.

VI. Tartışma

Çalışmanın ilk bulguları, üç temel sosyolojik örüntüye işaret etmektedir. Birincisi, teknolojik yayılım ile kültürel uyum arasında belirgin bir mesafe ortaya çıkmıştır. Bulgular bölümünde ortaya konan yaygın kullanım oranları ile kurumsal düzeydeki uyum eksiklikleri arasındaki açıklık, bu mesafenin somut göstergesi olarak değerlendirilebilir. Söz konusu örüntü, Ogburn'ün kültürel gecikme kavramıyla örtüşmektedir. Ogburn'e göre maddi kültür teknolojik yeniliklerle hızla değişirken, değerleri, normları ve kurumları kapsayan maddi olmayan kültür bu değişime daha yavaş uyum sağlamakta ve aradaki uyumsuzluk dönemi toplumsal sorunların kaynağı haline gelmektedir [10]. Sonraki çalışmalar da kültürel gecikmenin, kurumsal uyumun gecikmesini açıklayan temel çözümleme araçlarından biri olduğunu vurgulamaktadır [11]. Bununla birlikte literatürde kültürel gecikme kavramına yönelik eleştirel okumalar da mevcuttur; bu eleştiriler, kavramın maddi/maddi olmayan kültür ayrımını fazlasıyla ayrıştırıcı kurduğunu ve teknoloji-toplum ilişkisini tek yönlü bir nedensellik üzerinden okuma eğilimi taşıdığını ileri sürmektedir. Bu sınırlılık göz önüne alındığında; araştırma, Türk dünyasında teknolojik araçların kullanımı ile bunların sorumlu kullanımına ilişkin kurumsal kapasite arasında bir gecikme bulunduğunu göstermektedir. Bu örüntü Ogburn'cü çerçeveyle uyumlu olmakla birlikte; Türk dünyasında aynı anda gözlemlenen yüksek kullanım ve yüksek kararsızlık eş zamanlılığı, kurumsal gecikme ile bireysel düzeydeki ikircikli benimsemenin birbirini beslediği daha karmaşık bir dinamiğe işaret etmekte ve klasik kültürel gecikme okumasının tek başına yeterli olmayabileceğini düşündürmektedir.

İkincisi, dijital egemenlik ortak bir kaygı alanı olarak öne çıkmıştır. Beş ülkede de küresel teknoloji tekellerine, dış bulut altyapılarına, yabancı yapay zeka modellerine ve dış veri ekosistemlerine bağımlılık, stratejik bir kırılganlık olarak değerlendirilmiştir. Bu bulgu, dijital egemenliği devletlerin, şirketlerin ve yurttaşların dijital teknolojiler üzerinde denetim kurma çabası olarak tanımlayan araştırma ile örtüşmektedir [12]. İlgili çalışma, büyük teknoloji şirketlerinin yapay zeka üzerinde yüksek düzeyde fiili denetim kurduğunu ve bu durumun ulusal egemenlik açısından meşruiyet sorunları doğurduğunu ileri sürmektedir [13]. Öte yandan literatürde dijital egemenlik söyleminin aşırı ulus-devlet merkezli okunmasına, küresel açık bilim ve teknoloji iş birliği olanaklarını daraltma potansiyeline ilişkin eleştirel okumalar da mevcuttur; bu okumalar, egemenlik vurgusunun dijital korumacılıkla örtüşme riskine işaret etmektedir. Araştırmanın bulguları, Türk dünyasında ortak veri yönetişimi ve bölgesel dil kaynaklarına dayalı yapay zeka kapasitesi geliştirme fikrinin önem kazandığını ortaya koymuştur. Bu bulgu, dijital egemenlik tartışmasıyla doğrudan ilişkili olmakla birlikte; korumacı-baskın bir egemenlik anlayışı ile küresel iş birliğine açık bir egemenlik anlayışı arasında nasıl konumlandığı ayrıca ele alınmalıdır.

Üçüncüsü, yapay zeka temelli dijital dönüşüm bölgesel iş birliğini öne çıkarmıştır. Bulgular, Türk devletlerinin ayrı ayrı ulusal stratejiler geliştirdiğini; buna karşılık ortak dil, ortak kültürel miras ve kısmen örtüşen veri alanlarının var olduğunu göstermiştir. Özellikle doğal dil işleme, konuşma tanıma ve makine çevirisi gibi dile bağımlı yapay zeka uygulamalarında Türk dillerinin yapısı ve söz dizimsel benzerlikleri, ortak dil modellerinin geliştirilmesine elverişli bir zemin sunmaktadır. Bununla birlikte, bu olanağın henüz sistematik bir bölgesel politikaya dönüşmediğini göstermektedir, ülkelerin yapay zeka stratejileri büyük ölçüde birbirinden bağımsız biçimde kurgulanmakta, ortak veri havuzları ya da paylaşımlı altyapı girişimleri sınırlı kalmaktadır. Bu parçalı yapı, kaynak kısıtlı ülkelerde teknik kapasitenin yeterince ölçeklenememesine ve benzer sorunlara karşı tekrarlayan çözüm maliyetlerine yol açma riski taşımaktadır.

Araştırmanın stratejik önem ve yatırım algısına ilişkin bulguları, yüksek bir kararsızlık düzeyine işaret etmektedir. Bu örüntü, yapay zeka yönetişimine ilişkin kamuoyu araştırmalarıyla benzerlik göstermektedir; ilgili çalışmalar, toplumların yapay zekayı önemli bulmakla birlikte bu teknolojiyi yönetecek kurumlara duydukları güvenin düşük ve orta düzeyde kaldığını, sonuçlara ilişkin belirsizliğin yaygın olduğunu ortaya koymaktadır [14]. Bu karşılaştırma, Türk dünyasında gözlemlenen kararsızlığın bölgeye özgü bir durum olmadığını, yapay zekaya yönelik küresel kamuoyu eğilimleriyle tutarlı olduğunu düşündürmektedir. Bütün olarak değerlendirildiğinde araştırmanın bulguları, yapay zekanın Türk dünyasındaki yayılımını salt teknik bir altyapı sorununa indirgeyen okumaların yetersiz kaldığını; meseleyi toplumsal, kültürel ve stratejik boyutlarını birlikte içeren bir çerçeveden ele almayı gerektirdiğini göstermiştir.

VII. Öneriler

Araştırmanın ilk bulgularından hareketle Türk dünyası için dört stratejik öneri geliştirilebilir. İlk olarak, Türk devletleri ortak bir Türk dilli büyük dil modeli geliştirmelidir. Bu öneri; araştırmanın dijital egemenlik bölümünde ortaya konan yabancı yapay zeka modellerine bağımlılık kaygısına ve beş ülke kamuoyunda bölgesel dil kaynaklarına dayalı kapasite ihtiyacına ilişkin bulgulara dayanmaktadır. Bu model, Azerbaycan Türkçesi, Kazakça, Kırgızca, Özbekçe, Türkiye Türkçesi ve diğer Türk lehçelerini kapsayan geniş bir veri birikimine dayanmalıdır. Modelin geliştirilmesi; dilsel çeşitliliğin dijital ortamda korunmasına, kültürel egemenliğin güçlenmesine ve dış modellere bağımlılığın azaltılmasına katkı sağlayacaktır. Bu yatırımın stratejik etkisi; teknik çıktı düzeyiyle birlikte Türk dünyasının küresel yapay zeka ekosisteminde üretici aktör olarak konumlanması açısından değerlendirilmelidir.

İkinci olarak, yapay zeka okuryazarlığı temel bir vatandaşlık becerisi olarak ele alınmalıdır. Bu öneri; araştırmada katılımcıların büyük bir çoğunluğunun yapay zekayı düzenli olarak kullandığını gösteren bulguya ve yapısal risk alanlarına (veri mahremiyeti, dezenformasyon, algoritmik yanlılık, platform bağımlılığı) doğrudan karşılık gelmektedir. Yaygın kullanımın yüksek olması; kullanıcıların bilinçli, güvenli ve etik kullanım kapasitelerinin aynı düzeyde geliştiği anlamına gelmemekte, okuryazarlık politikaları stratejik bir gereklilik haline gelmektedir. Eğitim kurumları, kamu kuruluşları ve sivil toplum örgütleri; sorumlu yapay zeka kullanımı, dezenformasyonla mücadele, veri mahremiyeti, algoritmik farkındalık ve etik kullanım konularında yaş gruplarına ve meslek alanlarına göre farklılaştırılmış programlar geliştirmelidir. Bu programların; örgün eğitim müfredatlarına entegrasyonu, yaşam boyu öğrenme platformları aracılığıyla yaygınlaştırılması ve kamu bilgilendirmeleri ile kültürel ürünler aracılığıyla genel nüfusa erişmesi önerilmektedir. Bu yaklaşım, kullanım yoğunluğu ile etik farkındalık arasındaki açıklığı kapatmak suretiyle; yapay zekanın toplumsal düzeyde oluşturduğu kırılganlıkları azaltmaya katkı sağlayacaktır.

Üçüncüsü, mesleki dönüşüm ve yaşam boyu öğrenme programları güçlendirilmelidir. Katılımcıların azımsanamayacak bir bölümünün yapay zekanın işsizlik riski konusunda kararsız kalmış; emek piyasasının üç eş zamanlı dönüşüm dinamiği (otomasyon, yeniden tanımlama, yeni iş alanları) bu gerekliliği pekiştirmektedir. Kararsızlık ne kadar uzun sürerse bireysel ve kurumsal düzeyde hazırlık maliyetinin o kadar arttığı dikkate alındığında; bu alandaki politika müdahalesi zamana duyarlı bir nitelik kazanmaktadır. Üniversiteler, meslek örgütleri ve kamu kurumları, yapay zeka çağının yeni becerilerine yönelik ortak eğitim programları hazırlamalıdır. Programların; mesleki dönüşüm haritaları üzerine kurulması, yeniden beceri kazanımı ve beceri yükseltme modellerini birlikte içermesi ve kırılgan işgücü gruplarına (düşük nitelikli iş kolları, kadınlar, kırsal nüfus, genç iş arayanlar) öncelikli erişim sağlaması önem taşımaktadır. Bu yaklaşım, mevcut kararsızlığı dönüşüme hazır bir bireysel ve kurumsal kapasiteye çevirmenin temel koşulu olarak öne çıkmaktadır.

Dördüncüsü, bölgesel yapay zeka yönetişimi kurumsallaştırılmalıdır. Ortak dil, ortak kültürel miras ve kısmen örtüşen veri alanları; bölgesel yönetişim mekanizmalarının, ulusal stratejilere alternatif oluşturmaksızın bu stratejileri tamamlayıcı bir üst çerçeve olarak kurgulanmasını mümkün kılmaktadır. Türk Devletleri Teşkilatı ve ilgili kurumlar bünyesinde; ortak etik standartlar, paylaşılan veri yönetişimi ilkeleri, bölgesel uzman ağları, ortak araştırma platformları ve politika izleme mekanizmaları oluşturulabilir. Bu yapılar; ülkeler arasında bilgi paylaşımını hızlandırmak suretiyle politika öğrenmesini desteklemekte; ulusal düzeyde tek tek yanıtlanması güç olan yönetişim sorunlarına ölçek ekonomisi sağlayacaktır. Sonuç itibarıyla yukarıda sıralanan dört öneri; Türk dünyasında yapay zekanın sürükleyici bir dönüştürücü güç olmaktan çıkıp yönetilebilir bir kapasiteye dönüşmesinin temel bileşenleri olarak değerlendirilmelidir.

VIII. Sonuç

Bu çalışma, Türk Barometresi 2026 projesinin ilk bulgularından hareketle Türk dünyasında yapay zeka temelli dijital dönüşümün toplumsal görünümünü değerlendirmeyi amaçlamıştır. Çalışmanın özgün katkısı, beş Türk devletini (Azerbaycan, Kazakistan, Kırgızistan, Özbekistan ve Türkiye) ortak bir karşılaştırmalı çerçevede ele alarak ampirik çözümlemeyi sunmasıdır. Bulgular; yapay zekanın Türk dünyasında katılımcıların düzenli olarak kullandığı gündelik bir teknolojiye dönüştüğünü ortaya koymuştur. Yapay zekaya yönelik baskın duygunun korku ya da reddediş yerine; merak ve heyecan olduğu belirlenmiştir. Bu örüntü; kamu politikaları, eğitim sistemleri ve toplumsal farkındalık programları için stratejik bir politika bakış açısı sunmuştur. Bununla birlikte istihdam belirsizliği, yatırım algısında yüksek düzeyde kararsızlık, ülkeler arası kapasite farklılaşması, dijital okuryazarlık eksiklikleri ve veri egemenliği kaygıları şeklinde beş temel kırılganlık alanı belirlenmiştir.

Türk dünyasının önündeki temel stratejik soru iki seçenek arasında yoğunlaşmaktadır; yapay zeka teknolojilerini dışarıdan transfer eden ve yalnızca kullanan bir bölge olarak kalmak ya da kendi diline, verisine, insan kaynağına ve kültürel birikimine dayalı üretici bir dijital gelecek inşa etmek. Çalışmada ele alınan dijital egemenlik ve kültürel gecikme çerçeveleriyle doğrudan ilişkili olan bu ikilem; yapay zekanın Türk dünyası açısından teknolojik bir araç olmanın ötesinde ortak gelecek tasavvurunun, kültürel sürekliliğin, ekonomik rekabetin ve stratejik bağımsızlığın temel bir bileşenine dönüştüğünü göstermektedir. Tarihsel olarak Türk dünyasının kültürel ve fikrî dayanışmasını ifade eden “dilde, fikirde, işte birlik” anlayışı da yapay zeka çağında yeni bir boyut kazanmakta ve “dilde, fikirde, işte ve dijitalleşmede birlik” biçiminde yeniden yorumlanarak bu üretici geleceğin kavramsal çerçevesini oluşturabilecektir.

Kaynakça

[1] Stanford Institute for Human-Centered Artificial Intelligence, AI Index Report 2025, Stanford University, 2025.

[2] OECD, OECD Artificial Intelligence Principles, Paris: OECD, 2019, updated 2024.

[3] UNESCO, Recommendation on the Ethics of Artificial Intelligence, Paris: UNESCO, 2021.

[4] World Economic Forum, The Future of Jobs Report 2025, Geneva: World Economic Forum, 2025.

[5] Türkiye Cumhuriyeti Cumhurbaşkanlığı Dijital Dönüşüm Ofisi ve Sanayi ve Teknoloji Bakanlığı, Ulusal Yapay Zeka Stratejisi 2021–2025, Ankara, 2021.

[6] Government of Kazakhstan, Concept for Artificial Intelligence Development for 2024–2029, Astana, 2024.

[7] Republic of Uzbekistan, Strategy for the Development of Artificial Intelligence Technologies until 2030, Tashkent, 2024.

[8] Ministry of Economy of the Republic of Azerbaijan, Artificial Intelligence Strategy of the Republic of Azerbaijan for 2025–2028, Baku, 2025.

[9] Kyrgyz Republic, Concept for the Digital Transformation of the Kyrgyz Republic for 2024–2028, Bishkek, 2024.

[10] W. F. Ogburn, Social Change with Respect to Culture and Original Nature. New York: B. W. Huebsch, 1922.

[11] R. L. Brinkman and J. E. Brinkman, “Cultural lag: conception and theory,” International Journal of Social Economics, vol. 24, no. 6, pp. 609–627, 1997, doi: 10.1108/03068299710179026.

[12] J. Pohle and T. Thiel, “Digital sovereignty,” Internet Policy Review, vol. 9, no. 4, 2020, doi: 10.14763/2020.4.1532.

[13] H. Roberts, “Digital sovereignty and artificial intelligence: a normative approach,” Ethics and Information Technology, vol. 26, no. 4, 2024, doi: 10.1007/s10676-024-09810-5.

[14] B. Zhang and A. Dafoe, “U.S. public opinion on the governance of artificial intelligence,” in Proc. AAAI/ACM Conf. AI, Ethics, and Society, 2020, pp. 187–193, doi: 10.1145/3375627.3375827.

[15] A. Agrawal, J. S. Gans, and A. Goldfarb, “Artificial intelligence: the ambiguous labor market impact of automating prediction,” Journal of Economic Perspectives, vol. 33, no. 2, pp. 31–50, 2019, doi: 10.1257/jep.33.2.31.

[16] S. Zuboff, “Big other: surveillance capitalism and the prospects of an information civilization,” Journal of Information Technology, vol. 30, no. 1, pp. 75–89, 2015, doi: 10.1057/jit.2015.5.