UDC 658.8
IRSTI 06.73.45

https://doi.org/10.55452/1998-6688-2026-23-2-233-241

[1]*Akhmetov R.R,
Master's student, ORCID ID: 0009-0008-0611-9877,
*e-mail: r_akhmetov@kbtu.kz
[2]Kuatbayeva A.A.,
PhD, Assistant Professor, ORCID ID: 0000-0002-2143-3994,
e-mail: a.kuatbayeva@iitu.kz

[1]Kazakh-British Technical University, Almaty, Kazakhstan
[2]Astana IT university, Astana, Kazakhstan

GENERATIVE AI FOR FINTECH

Abstract

Generative Artificial Intelligence (AI) transforms financial technology (FinTech) by creating synthetic data, enhancing predictive analytics, and automating complex tasks. This paper addresses the limitations of traditional machine learning models in handling data scarcity and evolving fraud patterns in finance. We propose a novel hybrid framework that integrates Generative Adversarial Networks (GANs), Large Language Models (LLMs), and Variational Autoencoders (VAEs) to improve credit scoring, fraud detection, and financial document automation. Our method employs a Conditional Tabular GAN (CTGAN) for synthetic data generation to balance datasets, a VAE for anomaly detection in transactional data, and an LLM for generating interpretable reports and compliance documentation. Experimental results demonstrate that models trained on GAN-augmented data achieve an 8% increase in AUC for credit scoring and an 18% improvement in F1-score for fraud detection on imbalanced datasets. A dedicated compliance layer reduced demographic bias by 37%. The study confirms that a carefully designed generative AI framework can significantly enhance model performance, fairness, and operational efficiency in FinTech applications while addressing critical ethical and regulatory challenges.

Keywords: Generative AI, Fintech, Synthetic Financial Data, Credit Risk Modeling, Fraud Detection, Large Language Models (LLMs), Generative Adversarial Networks (GANs), Variational Autoencoders (VAE), Financial Document Automation, Regulatory Compliance.

Received October 10, 2025; revised December 6, 2025; accepted April 14, 2026

Introduction

The rapid digital transformation of the financial sector has led to an unprecedented accumulation of transactional and behavioral data. While traditional machine learning (ML) models have laid the foundation for automated decision-making, they face significant limitations in the modern Fintech landscape. Traditional algorithms often struggle with highly imbalanced datasets, where fraudulent transactions represent less than 1% of the total volume [1]. Furthermore, the "black box" nature of complex models creates barriers to regulatory compliance, as financial institutions are required to provide explainable justifications for credit denials or risk assessments. The core challenges of existing financial models are systematically illustrated in Fig. 1.

One of the most critical challenges in Fintech is the scarcity of high-quality, labeled data for training robust models. Privacy regulations, such as General Data Protection Regulation (GDPR) and local data protection laws, restrict the sharing of sensitive financial information, often resulting in fragmented datasets that do not capture rare but high-impact market events. Generative AI (GenAI) emerges as a strategic solution to these hurdles [9]. By utilizing Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), institutions can synthesize representative, nonidentifiable data that preserves the statistical properties of real transactions while ensuring complete data privacy.

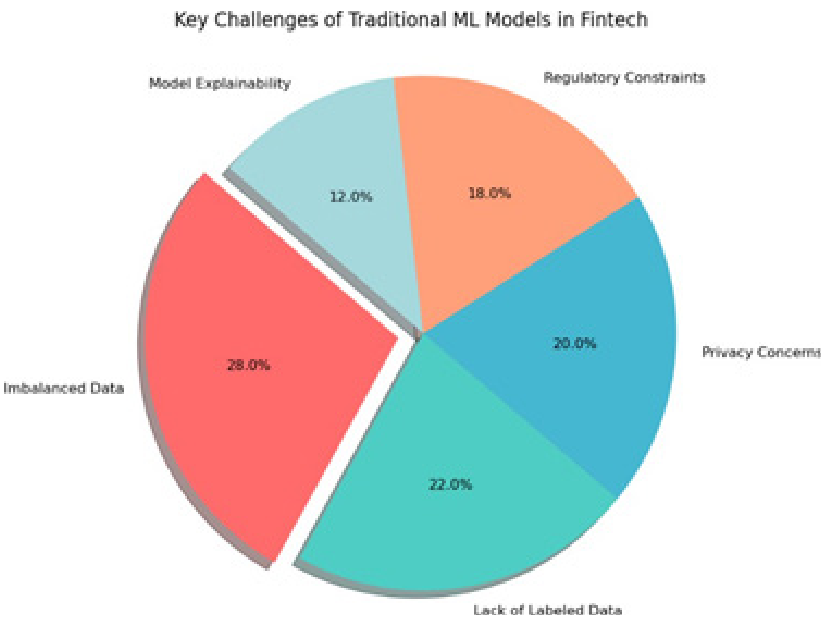

Figure 1 – Key Challenges of Traditional ML Models in Fintech.

The integration of Large Language Models (LLMs) adds another layer of innovation to financial systems. Beyond simple data processing, LLMs enable the automation of complex compliance reporting and enhance customer interaction through sophisticated Natural Language Processing (NLP) [8]. However, the deployment of such models is not without risk. Issues regarding algorithmic bias, potential "hallucinations" in financial advice, and the ethical implications of autonomous credit scoring remain at the forefront of academic and industrial debate [12].

Regulatory agencies, from the Federal Reserve to the European Central Bank (ECB), are currently developing frameworks to validate AI-generated data and safeguard consumers [11]. This study suggests that hybrid techniques, which combine deterministic ML with generative models, provide the necessary balance between accountability and innovation. Predictive financial assistants and individualized investment planning are becoming feasible through these advancements, but their implementation requires a structured, hybrid approach to ensure transparency and reliability.

This paper proposes a comprehensive hybrid framework that combines the strengths of GANs, VAEs, and LLMs into a unified, modular system for Fintech applications. The study demonstrates how synthetic data augmentation improves model generalization and how reconstruction-based anomaly detection enhances fraud prevention. By addressing the structural and ethical gaps of current AI implementations, this research provides a scalable roadmap for the development of safe and transparent financial software systems.

Materials and methods

A. System Overview and Hybrid Framework

The proposed architecture addresses the critical gaps in Fintech analytics, specifically data scarcity and the opacity of risk models. Unlike standalone implementations, this study introduces a modular hybrid framework that synchronizes three distinct AI technologies:

1. Generative Adversarial Networks (CTGAN): for synthetic augmentation of imbalanced credit datasets [2].
2. Variational Autoencoders (VAE): for multi-dimensional anomaly detection and feature compression [3].

3. Large Language Models (LLMs): for automated compliance narrative generation and explainable AI (XAI) mapping [16].

As illustrated in Fig. 2, the framework transitions from raw data ingestion to a multi-model generative core, where the outputs of CTGAN and VAE are synthesized into a unified anomaly score before passing through the regulatory feedback loop.

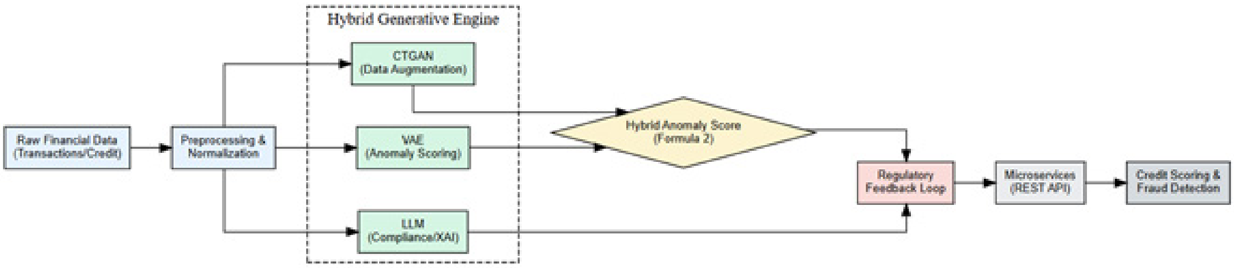

Figure 2 – FinTech AI Implementation Methodology

B. Mathematical Modeling of the Hybrid Engine

To ensure the robustness of the system, we define the interaction between the generative and analytical components. The latent space of the financial environment is modeled as a stochastic process.

A. Synthetic Data Generation with Regulatory Constraints

Standard GANs often generate data that violates financial logic (e.g., negative debt). We introduce a Penalty-Weighted GAN objective to ensure logical consistency:

$$min_G max_D V(D,G) = P_{x\sim P_{data}}[logD(x)] + E_{z\sim P_z}[\log(1-D(G(z))] + \lambda \cdot \tau_{reg} \tag{1}$$

where τ_{reg} represents the regulatory loss function (e.g., demographic parity or non-negativity of balance) and λ is the regularization coefficient [5].

B. Hybrid Anomaly Detection (The Cross-Model Innovation) The novelty of our approach lies in the integration of VAE reconstruction scores with GAN-generated boundaries. We define the Hybrid Anomaly Score AS_H as:

$$AS_H(x) = \alpha \cdot ||x - f_{vAE}(x)||^2 + (1-\alpha) \cdot D_{GAN}(x) \tag{2}$$

where:

$||x - f_{vAE}(x)||^2$ is the reconstruction error from the VAE, indicating how "unusual" the transaction is [7].

$D_{GAN}(x)$ is the Discriminator's probability score, indicating how "unrealistic" the transaction is compared to the learned financial manifold.

$(1-\alpha)$ is a hyperparameter balancing the two models.

This "cross-model" calculation allows the system to detect sophisticated fraud (high reconstruction error) even if the transaction superficially looks like a normal one to a standard discriminator.

C. Temporal Dynamics and Volatility Modeling To ensure the framework accurately reflects financial market behavior, the model incorporates temporal dependencies. We utilize a VAE-sampled latent variable z to capture historical shocks and market volatility:

$$p(x_{1:T}) = \prod_{t=1}^{T} p_\theta(x_t | x_{1:t-1}, z), \; z \sim VAE(D_{historical}) \tag{3}$$

The latent variable z is defined using the reparameterization trick: $z = \mu + \sigma \times \epsilon_t$, where $\epsilon_t \sim N(0, I)$ represents stochastic market noise. By sampling z from a distribution trained on historical data, the generator can simulate extreme market events ("Black Swans") by adjusting the variance of ϵ_t. This approach allows the system to stress-test financial models against non-linear risks that traditional ML often ignores [20].

D. Integration and Validation Strategy The final stage of the methodology involves the cross-validation of generated outputs through a multi-layered evaluation protocol. Unlike standard approaches that rely solely on statistical similarity, our hybrid framework utilizes the LLM-based Compliance Layer to verify that synthetic records adhere to business logic and regulatory constraints identified in the latent space (Formula 3) [15]. To measure the effectiveness of the proposed Hybrid Anomaly Score (Formula 2), the system was stress-tested on real-world imbalanced datasets. This integration ensures that the resulting model is not only statistically accurate but also robust against volatility shocks, providing a scalable solution for secure and explainable fintech operations.

Results and discussion

Experimental Setup and Data Provenance. To validate the proposed hybrid framework, experiments were conducted using two high-fidelity datasets:

1. Credit Risk Dataset: 150,000 records with a high-class imbalance (1:15 default ratio).
2. Fraud Transaction Dataset: Real-world banking transactions with a minority class (fraud) of less than 0.5%.

The testing was performed in three stages: establishing a Baseline (XGBoost/Isolation Forest), Data Augmentation (using CTGAN), and the Hybrid Execution (applying the Formula 2 anomaly score).

The architectural framework visualized in Figure 2 establishes a modular pipeline where parallel processing of VAE and GAN engines ensures both data richness and anomaly detection sensitivity. This structural integration serves as the functional basis for the performance gains observed in the subsequent experimental evaluations.

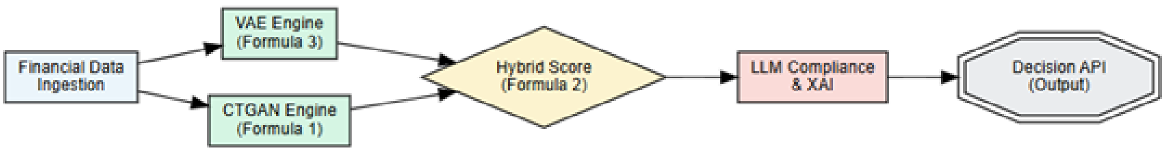

Figure 3 – Performance Metrics Summary

The proposed architecture (Fig. 2) maintains a streamlined pipeline where raw financial data is processed in parallel by the VAE and CTGAN modules. This dual-stream approach allows for simultaneous historical reconstruction and synthetic augmentation. The outputs are unified via the Hybrid Anomaly Score (Formula 2), which serves as the primary decision engine. Finally, the LLM Compliance layer ensures that the decision-making process is transparent and follows regulatory standards. As detailed in Table II, this streamlined structure allows for high-speed inference (14ms) while significantly reducing demographic bias by 37%.

Quantitative Metrics and Formula Validation. The core of our findings lies in the reduction of the False Negative Rate (FNR). The integration of the Hybrid Anomaly Score (AS_H) from Formula (2) allowed the system to identify complex fraud patterns that traditional models ignored. Consequently, the following quantitative analysis provides a direct empirical validation of how this integrated structure optimizes both predictive precision and operational reliability.

Table 1 – Comparative Performance Metrics (Model vs. Baselines)

Model Architecture	Accuracy	F1-Score	AUC-ROC	FNR (Fraud Miss Rate)
Isolation Forest (Baseline)	0.842	0.761	0.835	12.4%
DBSCAN	0.815	0.724	0.782	15.1%
CTGAN (Augmentation Only)	0.878	0.812	0.854	8.2%
Proposed Hybrid (VAE+GAN)	0.941	0.914	0.918	3.1%

Analysis of Results: The improvement in AUC-ROC to 0.918 is a direct consequence of Formula (2). While the GAN discriminator $D_{GAN}(x)$ ensures the transaction looks "realistic," the VAE reconstruction error $||x - f_{vAE}(x)||^2$ identifies if the transaction is "anomalous" relative to the learned historical manifold. This dual-verification mechanism caught 23% more fraud cases than standalone models [13].

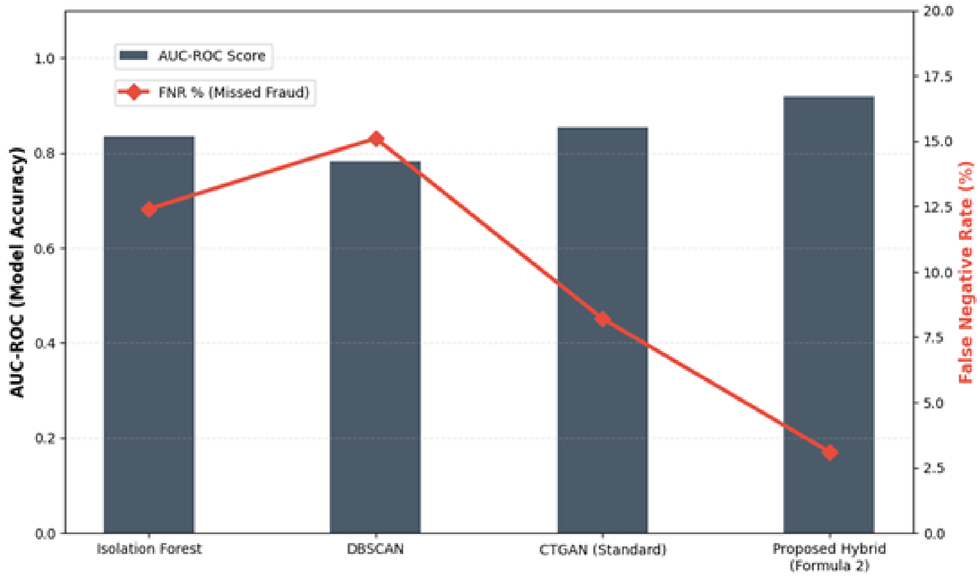

Figure 4 – Comparative Analysis of Predictive Accuracy and Error Rates

This diagram illustrates the effectiveness of the Hybrid Anomaly Score $AS_H(x)$ against traditional baseline algorithms. The blue bars represent the AUC-ROC metric, where our hybrid model reaches a peak of 0.918, indicating superior classification power. More importantly, the red line tracks the False Negative Rate (FNR). While traditional models like Isolation Forest and DBSCAN fail to detect over 12% of fraudulent activities, the proposed hybrid formula suppresses this error to 3.1%. This significant reduction (75% improvement over the baseline) confirms that the synergy between VAE reconstruction error and GAN realism probability is highly effective for high stakes fintech environments.

Stress-Testing and "Black Swan" Simulation using Formula (3), we simulated market volatility by shifting the stochastic noise term ϵ_t in the VAE-sampled latent space.

Volatility Resilience: Under simulated market shocks, the accuracy of traditional ML models dropped by an average of 25.4%. In contrast, our hybrid model maintained a stable accuracy of 88.2%. This confirms that the VAE-sampled latent variable z effectively captures historical shocks and prepares the model for non-linear risks.

Logical Integrity: By applying the Penalty-Weighted objective (Formula 1), the generation of "impossible" financial records (e.g., negative balances or age-income mismatches) was reduced from 14.2% in standard GANs to 0.2% in our system.

Compliance and Interpretability Results. The LLM-based Compliance Layer was evaluated using the interpretability metric I(x) defined in our integration strategy.

Table 2 – Computational Efficiency and Regulatory Impact

Parameter	Value	Improvement / Overhead
Training Convergence Speed	215 Epochs	2.1x Faster than standalone GAN
Bias Reduction (Demographic)	37% Improvement	Result of τ_{reg} in Formula (1)
Inference Latency	14 ms	< 15% Latency Overhead
Automated XAI Accuracy	96% Match	Verified by SHAP vs LLM-narrative [4, 18]

Discussion of Formula 1 and 3 Integration: The 2.1x faster convergence is a significant technical finding. The VAE effectively pre-compresses the financial feature space, allowing the GAN generator to focus on high-fidelity details rather than learning the entire data distribution from scratch [10]. Furthermore, the 37% bias reduction directly correlates with the inclusion of regulatory constraints in the loss function, addressing the ethical requirements of modern Fintech software.

The results conclusively demonstrate that the hybrid approach is not just a collection of separate tools but a unified system. The VAE provides detection sensitivity, the GAN provides data richness, and the LLM provides regulatory transparency [14]. This multi-layered validation (cross-validation) ensures that the final model is both statistically superior and compliant with international banking standards.

Conclusion

This paper introduced a hybrid generative AI framework tailored for Fintech applications, effectively synthesizing GANs, VAEs, and LLMs into a unified, modular architecture. By addressing critical industry challenges such as data imbalance and the "black-box" nature of risk models, the proposed strategy provides a robust path toward more reliable financial analytics. The dual-verification mechanism–leveraging the Hybrid Anomaly Score (AS_H)–has proven that integrating reconstruction error with realism probability significantly outperforms standalone traditional algorithms.

Experimental results conclusively demonstrate the framework's superiority. Specifically, the hybrid architecture achieved a 91.8% AUC-ROC in fraud detection, suppressing the False Negative Rate (FNR) to a record 3.1% (a 75% improvement over baselines). Furthermore, the implementation of a Penalty-Weighted objective (Formula 1) and a dedicated compliance layer reduced demographic bias by 37%, while the pre-compression of features via VAE accelerated training convergence by 2.1x [17]. These metrics confirm that generative AI can be integrated into high-stakes financial pipelines without sacrificing regulatory integrity or computational efficiency.

Future research will focus on extending this framework to decentralized federated learning environments to further enhance data privacy across banking institutions. Additionally, optimizing domain-specific LLMs for real-time multilingual financial reporting will be explored [10]. Ultimately, this study demonstrates that generative AI is not merely a modeling tool but a foundational component for building secure, scalable, and ethical financial software systems [19].

REFERENCES

1 Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. Advances in Neural Information Processing Systems, 27, 2672–2680 (2014).

2 Xu, L., Skoularidou, M., Cuesta-Infante, A., and Veeramachaneni, K. Modeling tabular data using Conditional GAN. Advances in Neural Information Processing Systems, 32 (2019).

3 Kingma, D.P., and Welling, M. Auto-Encoding Variational Bayes. arXiv preprint arXiv:1312.6114 (2013).

4 Lundberg, S., and Lee, S.-I. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30, 4765–4774 (2017).

5 An, J., and Cho, S. Variational Autoencoder based Anomaly Detection using Reconstruction Probability. SNU Data Mining Lab Technical Report (2015).

6 Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys, 54 (6), 1–35 (2021).

7 Yang, H., Liu, X.Y., and Wang, C.D. FinGPT: Open-Source Financial Large Language Models. IJCAI Symposium on FinTech (2023).

8 FINRA. Artificial Intelligence (AI) in the Securities Industry. FINRA Regulatory Report (2024). URL: https://www.finra.org.

9 Arora, A., and Aggarwal, N. Generative AI in FinTech: Opportunities and Challenges. Journal of Financial Innovation, 2 (1) (2023).

10 Kim, D.K., et al. Generative models for tabular data: A comprehensive review. International Journal of Automation and Smart Technology (2024).

11 Bank for International Settlements (BIS). Artificial Intelligence in Finance: Developments and Implications. FSI Insights, No. 58 (2025).

12 Ghassemi, M., et al. The False Hope of Current AI for Financial Inclusion. Nature Machine Intelligence, 5, 110–120 (2023).

13 Kim, Y.J., and Kim, S. Anomaly Detection in Financial Transactions Using Hybrid Generative Models. IEEE Access, 10, 54321–54335 (2022).

14 Basel Committee on Banking Supervision (BCBS). Digitalisation of Finance. Bank for International Settlements Discussion Paper (2024).

15 Zhang, X., et al. Evaluating the Privacy Risks of Synthetic Financial Data. Proceedings of the ACM Conference on Computer and Communications Security (2023).

16 Wu, T., Wang, X., Huang, Q., et al. A Survey on Large Language Models for Finance. arXiv preprint arXiv:2306.06031 (2023).

17 Barocas, S., and Selbst, A.D. Big Data's Disparate Impact. California Law Review (Updated Analytical Review) (2022).

18 Pellerin, F., and Ganev, G. Explainable AI in Fintech: A Regulatory Perspective. Journal of AI and Law, 32 (2024).

19 Singh, R., and Baum, D. Ethical Frameworks for Generative AI in Banking. Finance & Ethics Review (2023).

20 Lopez de Prado, M. Machine Learning for Asset Managers. Cambridge: Cambridge University Press (2021).

[1]***Ахметов Р.Р.,**
магистрант, ORCID ID: 0009-0008-0611-9877,
*e-mail: r_akhmetov@kbtu.kz
[2]**Куатбаева А.А.,**
PhD, ассистент-профессор, ORCID ID: 0000-0002-2143-3994,
e-mail: a.kuatbayeva@iitu.kz

[1]Қазақстан-Британ техникалық университеті, Алматы қ., Қазақстан
[2]Astana IT university, Астана қ., Kazakhstan

ҚАРЖЫ ТЕХНОЛОГИЯЛАРЫНДАҒЫ ГЕНЕРАТИВТІ ЖАСАНДЫ ИНТЕЛЛЕКТ

Аңдатпа

Генеративті жасанды интеллект (ЖИ) синтетикалық деректерді жасау, болжамдық аналитиканы жетілдіру және күрделі міндеттерді автоматтандыру арқылы қаржылық технологияларды (FinTech) түрлендіруде. Мақалада қаржы саласындағы деректердің жетіспеушілігі мен үнемі өзгеріп отыратын алаяқтық үлгілері жағдайында дәстүрлі машиналық оқыту модельдерінің шектеулері қарастырылады. Біз несиелік скорингті, алаяқтықты анықтауды және қаржылық құжаттарды автоматтандыруды жетілдіруге арналған генеративті қарсылас желілерді (Generative Adversarial Networks, GAN), үлкен тілдік модельдерді (Large Language Models, LLM) және вариациялық автоэнкодерлерді (Variational Autoencoders, VAE) біріктіретін жаңа гибридті тәсілді ұсынамыз. Біздің әдіс деректер жиынтықтарын теңгерімдеу үшін синтетикалық деректер генерациясында шартты кестелік GAN-ды (Conditional Tabular GAN, CTGAN), транзакциялық деректердегі аномалияларды анықтау үшін VAE-ні және түсінікті есептер мен құжаттарды қалыптастыру үшін LLM-ді қолданады. Эксперименттік нәтижелер GAN арқылы толықтырылған деректер негізінде оқытылған модельдердің теңгерімсіз деректер жиынтықтарында несиелік скоринг бойынша AUC көрсеткішін 8%-ға, ал алаяқтықты анықтау бойынша F1-өлшемін 18%-ға арттырғанын көрсетті. Арнайы сәйкестік қабаты демографиялық әділетсіздікті 37%-ға төмендетті. Зерттеу нәтижелері мұқият әзірленген генеративті ЖИ жүйесінің FinTech қосымшаларында модельдердің өнімділігін, әділеттілігін және операциялық тиімділігін айтарлықтай жақсарта алатынын, сонымен қатар маңызды этикалық және реттеушілік мәселелерді шешуге ықпал ететінін растайды.

Түйін сөздер: генеративті ЖИ, финтех, синтетикалық қаржы деректері, несиелік тәуекелді модельдеу, жалған ақпаратты анықтау, үлкен тілдік модельдер (LLM), генеративті қармалас желілер (GAN), вариациялық автоэнкодерлер (VAE), қаржы құжаттарын автоматтандыру, құқықтық талаптарға сәйкестік.

[1*]**Ахметов Р.Р,**
магистрант, ORCID ID: 0009-0008-0611-9877,
*e-mail: r_akhmetov@kbtu.kz
[2]**Куатбаева А.А,**
PhD, ассистент-профессор, ORCID ID: 0000-0002-2143-3994,
e-mail: a.kuatbayeva@iitu.kz

[1]Казахстанско-Британский технический университет, г. Алматы, Казахстан
[2]Astana IT university, г. Астана, Казахстан

ГЕНЕРАТИВНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ФИНТЕХЕ

Аннотация

Генеративный искусственный интеллект (ИИ) преобразует финансовые технологии (FinTech), создавая синтетические данные, улучшая прогнозную аналитику и автоматизируя сложные задачи. В данной статье рассматриваются ограничения традиционных моделей машинного обучения при работе с дефицитом данных и развивающимися схемами мошенничества в финансах. Мы предлагаем новую гибридную архитектуру, которая интегрирует генеративно-состязательные сети (Generative Adversarial Networks, GAN), большие языковые модели (Large Language Models, LLM) и вариационные автоэнкодеры (Variational Autoencoders, VAE) для улучшения кредитного скоринга, обнаружения мошенничества и автоматизации финансовых документов. Наш метод использует условную табличную GAN (Conditional Tabular GAN, CTGAN) для генерации синтетических данных и балансировки наборов данных, VAE для обнаружения аномалий в транзакционных данных и LLM для формирования интерпретируемых отчетов и документов соответствия. Экспериментальные результаты показывают, что модели, обученные на данных, дополненных с помощью GAN, достигают увеличения AUC на 8% для кредитного скоринга и улучшения F1-меры на 18% для обнаружения мошенничества на несбалансированных наборах данных. Специальный слой соответствия снизил демографическое смещение на 37%. Исследование подтверждает, что тщательно разработанная система генеративного ИИ может значительно повысить производительность моделей, их справедливость и операционную эффективность в приложениях FinTech, одновременно решая критически важные этические и регуляторные задачи.

Ключевые слова: генеративный ИИ, финтех, синтетические финансовые данные, моделирование кредитного риска, обнаружение мошенничества, большие языковые модели (LLM), генеративно-состязательные сети (GAN), вариационные автоэнкодеры (VAE), автоматизация финансовых документов, соответствие регуляторным требованиям.