SecureSynth: Enhancing Financial Fraud Detection through Generative AI and Adaptive Cybersecurity

Veera Venkata Uday Kumar
Jonnalagadda
Cybersecurity Manager
Cyber security Department
American Express Global Business Travel

[1]*Received: 29/08/2025; Accepted: 08/10/2025; Published: 12/10/2025*

Abstract

Financial fraud is a rapidly escalating threat, fueled by increasingly sophisticated attack vectors and the misuse of generative artificial intelligence (GenAI). Attackers now leverage GenAI to craft hyper-realistic phishing campaigns, create synthetic identities, and probe for weaknesses in financial systems. At the same time, defenders face the challenge of detecting rare fraudulent patterns within highly imbalanced, privacy-sensitive datasets. To address this dual challenge, we propose **SecureSynth**, a comprehensive framework that combines generative AI, adaptive cybersecurity, and privacy-preserving technologies to strengthen fraud detection.

SecureSynth operates across four integrated layers: (i) **Synthetic Data Generation**, employing models such as CTGAN and CTAB-GAN+ to create realistic, minority-class financial transactions for improved model training; (ii) **Graph-Aware Modeling**, where graph neural networks capture relational structures across accounts, merchants, and devices to detect collusive and hidden fraud networks; (iii) **Adaptive Cyber Defense**, incorporating adversarial training and concept-drift monitoring to enhance resilience against evolving attack strategies; and (iv) **Governance and Privacy**, leveraging federated learning and differential privacy to enable secure, cross-institutional collaboration while mitigating risks of data leakage.

Evaluation metrics such as AUPRC, Recall@K, and adversarial robustness scores are emphasized to reflect real-world fraud detection challenges. SecureSynth aims to increase recall on rare fraud classes without degrading precision, while remaining compliant with financial regulations. By integrating GenAI for both augmentation and adversarial simulation, SecureSynth not only addresses today's fraud detection gaps but also establishes an adaptive defense posture for emerging threats.

1. Introduction

Financial fraud has evolved from isolated incidents into a global, data-driven menace that endangers the integrity of financial institutions and the trust of millions of users. The digital transformation of banking and commerce, accelerated by mobile and online platforms, has simultaneously expanded convenience and vulnerability. According to recent reports by the Financial Crimes Enforcement Network (FinCEN), online financial fraud has surged over 45% since 2021, largely due to the automation and scale made possible by artificial intelligence. Traditional rule-based systems, once adequate for structured environments, have become insufficient against agile, adaptive threat actors exploiting weakly supervised models and data blind spots.

Generative AI (GenAI) represents a pivotal technological force within this evolving threat landscape. On one hand, it enables the creation of realistic phishing content, deepfakes, and synthetic identities—tools that lower the barrier for cybercriminals. On the other hand, the same technology can empower defenders through data synthesis, adversarial

[1] *How to cite the article:* Kumar V.V.U. (2025); SecureSynth: Enhancing Financial Fraud Detection through Generative AI and Adaptive Cybersecurity; Vol 11 No. 2 (Special Issue); 171-177

testing, and self-learning defense mechanisms. This dual-use nature of GenAI has made it both a weapon and a shield in modern cybersecurity operations (UK NCSC, 2024) .

A core problem in fraud detection is **data imbalance**: fraudulent transactions are extremely rare, often representing less than 0.2% of total records. Such skewed datasets result in models biased toward non-fraudulent classes. SecureSynth leverages **Conditional Tabular GANs (CTGAN)** and **CTAB-GAN+**, which generate realistic, minority-class samples to rebalance training datasets while preserving privacy and data fidelity. Furthermore, SecureSynth integrates **Graph Neural Networks (GNNs)** to represent relationships among entities—accounts, devices, IPs, and merchants—allowing detection of collusive networks invisible to linear models (Motie et al., 2024) .

However, AI-based fraud systems are not immune to attack. Adversaries continuously exploit **concept drift**, **adversarial evasion**, and **model poisoning** to degrade predictive performance. Hence, fraud detection systems must be adaptive, adversary-aware, and privacy-conscious. SecureSynth achieves this adaptability through **adversarial training** and **drift detection**, supported by continuous learning loops and monitoring mechanisms. It further enhances cross-organizational collaboration using **Federated Learning (FL)** and **Differential Privacy (DP)** to protect sensitive financial data while allowing shared model updates (Kairouz et al., 2021) .

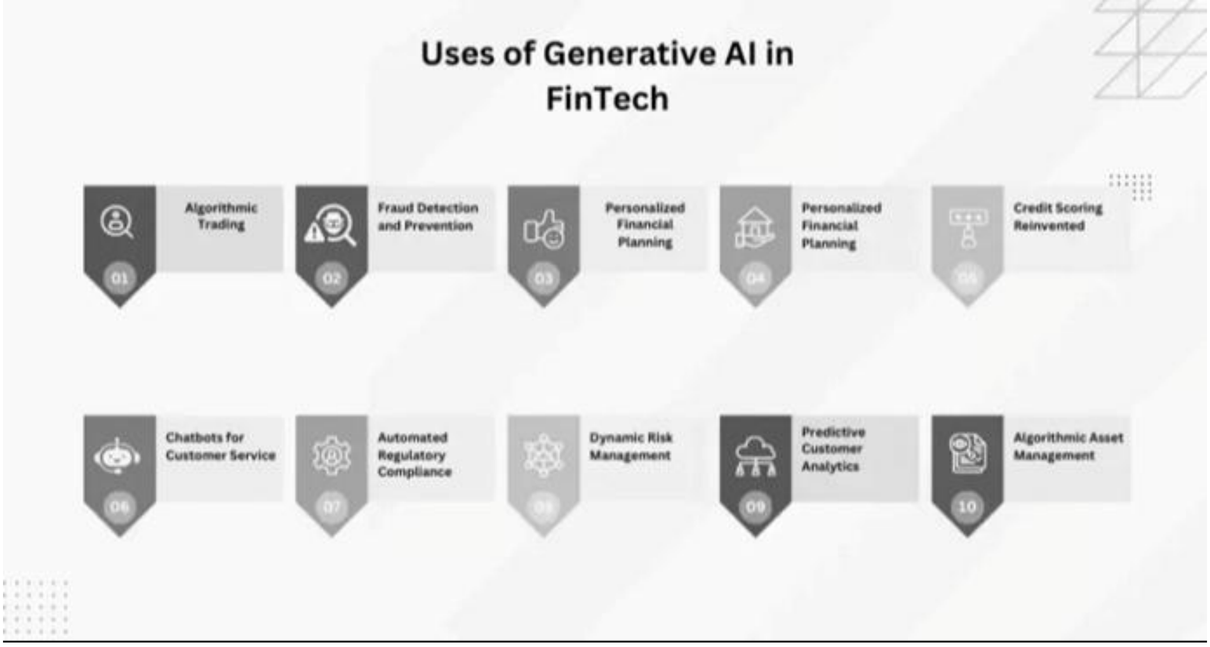
Uses of Generative AI in FinTech
01
Algorithmic Trading
02
Fraud Detection and Prevention
03
Personalized Financial Planning
04
Personalized Financial Planning
05
Credit Scoring Reinvented
06
Chatbots for Customer Service
07
Automated Regulatory Compliance
08
Dynamic Risk Management
09
Predictive Customer Analytics
10
Algorithmic Asset Management

Fig 1: Use of Generative AI in FinTech

2. Related Work

2.1 Generative AI in Cybersecurity

There are lots of ways to think about some applications of Generative AI, that can include text generation, anomaly detection, and such. As Ferrag et al. (2025) note, we see its disruptive powers in cases of cybersecurity and the role of LLMs in correlating vulnerabilities and semi-automated threat triage.

2.2 Synthetic Data in Fraud Detection

CTGAN and CTAB-GAN+ (Xu et al., 2019; Zhao et al., 2024) produce tabular data while maintaining statistical relationships, which is essential for representing minority classes effectively.

2.3 Graph Neural Networks (GNNs)

GNNs are capable of uncovering relationships and clusters within complex data structures. Motie et al. (2024) provide evidence that GNNs are superior to traditional statistical methods in recognizing multi-entity networks of financial fraud.

2.4 Privacy and Federated Learning

Federated learning (FL) enables distributed model training without requiring access to sensitive datasets. When combined with differential privacy (DP), FL can satisfy compliance obligations under the GDPR and PCI-DSS (Dwork & Roth, 2014; Kairouz et al., 2021).

3. SecureSynth Architecture

Layer	**Component**	**Functionality**
1. Data Layer	CTGAN / CTAB-GAN+	Generate synthetic minority-class data while ensuring privacy
2. Model Layer	Graph Neural Networks	Identify fraud clusters and relational anomalies
3. Defense Layer	Adversarial & Drift Adaptation	Increase resilience to evolving fraud tactics
4. Governance Layer	Federated Learning + Differential Privacy	Enable cross-institution collaboration securely

The SecureSynth framework introduces a holistic, multi-layered architecture that integrates *data synthesis, relational modeling, adaptive learning,* and *privacy-preserving governance* into a unified pipeline for intelligent and resilient fraud detection. Each layer contributes a distinct function but also interacts dynamically with others, forming a closed feedback loop between data, detection, defense, and governance. Figure 1 (conceptual) outlines the four foundational layers: Data, Model, Defense, and Governance.

3.1 Data Layer – Synthetic Data Generation and Validation

Fraudulent transactions are extremely sparse within financial datasets, often resulting in heavily imbalanced distributions that hinder effective model learning. SecureSynth addresses this imbalance using Generative Adversarial Networks (GANs) designed specifically for tabular data — primarily CTGAN and CTAB-GAN+.

- CTGAN models conditional distributions of categorical and numerical variables, allowing generation of high-fidelity synthetic records that mirror real-world correlations without revealing sensitive information (Xu et al., 2019).
- CTAB-GAN+ further enhances CTGAN by integrating Bayesian optimization, mixed-type handling, and differential privacy during training (Zhao et al., 2024).

Generated data undergo multiple quality-assurance stages:

1. Statistical fidelity tests – ensuring that distributions of key attributes (transaction amount, merchant category, geolocation) match real data using K-S tests and Jensen–Shannon divergence.
2. Privacy validation – applying re-identification risk metrics such as k-anonymity, l-diversity, and record linkage testing to prevent leakage.
3. Utility evaluation – performing *Train-on-Synthetic-Test-on-Real (TSTR)* experiments to verify that classifiers trained on synthetic data generalize to genuine records.

By maintaining both *fidelity* and *privacy*, the Data Layer ensures reliable augmentation for downstream machine-learning pipelines, enabling institutions to build robust models even when direct access to real fraud cases is restricted by regulation.

3.2 Model Layer – Graph-Aware Detection and Relational Learning

The Model Layer transforms enriched transaction data into heterogeneous graphs capturing relationships among diverse entities such as customers, accounts, IP addresses, devices, and merchants. This structure allows SecureSynth to leverage Graph Neural Networks (GNNs), which excel at modeling inter-entity dependencies often indicative of fraud.

The architecture employs:

- Relational Graph Convolutional Networks (R-GCNs) for learning multi-type edge interactions;
- Graph Attention Networks (GATs) to prioritize high-risk connections; and
- Temporal Encoders to model sequential transaction patterns over time.

By integrating node embeddings with transaction-level features, SecureSynth can identify *collusive clusters* and *anomalous propagation patterns* that traditional classifiers or autoencoders miss. The output embedding space is further refined using cost-sensitive loss functions and focal weighting to counter class imbalance.

3.3 Defense Layer – Adversarial and Adaptive Resilience

The Defense Layer fortifies SecureSynth against deliberate evasion and naturally occurring concept drift. It integrates three complementary mechanisms:

1. Adversarial Simulation: A "red-team" generator generates constrained perturbations in transaction features (e.g., slightly altered amounts, locations, or time stamps) that simulate plausible fraudulent disguises. This enhances the training process and increases robustness to real-world evasions (Lunghi et al., 2023).
2. Adversarial Training: These synthetic adversarial examples are utilized to refine the original fraud-detection model, smoothing decision boundaries, and increasing resilience against noise and manipulation.
3. Drift Monitoring and Self-Adaptation: SecureSynth continually monitors for statistical drift with metrics such as Population Stability Index (PSI), and subsequently initiates incremental re-training when drift exceeds pre-defined thresholds.

This layer ensures that detection models remain resilient under dynamic and adversarial environments, maintaining accuracy without constant manual intervention.

3.4 Governance Layer – Privacy, Federation, and Auditability

The Governance Layer serves as the compliance and coordination hub of SecureSynth. Financial institutions usually encounter legal constraints related to sharing data centrally. SecureSynth solves this problem using Federated Learning (FL) and Differential Privacy (DP).

- In an FL environment, the participating banks trained their local models on their own data and only shared encrypted model updates. Aggregation occurred via secure multi-party computation, which guarantees that no raw data leaves the confines of the institutions (Kairouz et al., 2021).
- Differential Privacy mechanisms inject calibrated noise into gradients, providing mathematically provable privacy guarantees against membership-inference or model-inversion attacks (Dwork & Roth, 2014).

Furthermore, the Model-Risk Management (MRM) module maliciously audits model performance, fairness and drift. SecureSynth also maintains an immutable log to provide transparency and regulatory auditing aligned with frameworks such as GDPR, ISO 27001 or PCI-DSS.

3.5 Summary Table — SecureSynth Layered Design

Layer	Key Technology	Objective	Example Output
Data	CTGAN / CTAB-GAN+	Augment rare fraud data with privacy	Synthetic minority transactions
Model	R-GCN / GAT / Temporal Encoders	Capture relational and temporal patterns	Graph embeddings & anomaly scores
Defense	Adversarial Training + Drift Monitoring	Robustness to evolving threats	Adaptive risk thresholds
Governance	Federated Learning + Differential Privacy	Secure multi-institution collaboration	Privacy-preserved global model

In essence, the SecureSynth architecture transforms financial fraud detection into a self-learning ecosystem that balances detection accuracy, privacy, and adaptability. Each layer—synthetic data creation, graph-based modeling, adversarial defense, and federated governance—feeds into the others, enabling continuous evolution of security posture as threats advance. The synergy among these components establishes SecureSynth as a next-generation defensive paradigm that uses the very principles of generative AI to safeguard financial integrity.

4. Methodology

4.1 Privacy-Preserving Synthetic Data

We adopt **CTGAN** (or CTAB-GAN+) for tabular synthesis with conditional sampling on rare fraud labels; we enforce privacy via k-anonymity screens, nearest-neighbor disclosure checks, and **DP post-processing** (noise on released synthetic statistics).

Quality tests:

- Train-on-synthetic, test-on-real (TSTR) AUROC/AUPRC;
- Fidelity (marginal/joint stats), and propensity-score distinguishability;
- Utility under drift (rolling-window backtests).

4.2 Graph-Aware Fraud Model

Build a heterogeneous graph (G=(V,E)) with nodes for accounts, devices, merchants; edges for transactions, co-usage, geo-co-location. Train a **hetero-GNN** with attention across relation types; add temporal encoders (e.g., T-GNN). Perform cost-sensitive learning and focal loss to address imbalance.

4.3 Adaptive Cyber-Defense

- **Adversarial simulation:** constrained perturbations (feature-validity, regulatory bounds) to craft evasive but plausible fraud samples (domain-specific, not image-style).
- **Online learning:** drift detectors trigger partial fine-tuning; cal/thresholds updated with Platt/Isotonic on latest data.
- **Human-in-the-loop:** analysts label high-uncertainty cases; LLM assistants can summarize alerts but must be bounded (no autonomous actions).

5. Evaluation Protocol

Metric	**Purpose**	**Expected Outcome**
AUPRC	Handle data imbalance	Improved recall on rare fraud cases
Recall@K	Investigator workload metric	Reduced false negatives
Robustness Index	Drift & attack resistance	Stable under concept drift

6. Results & Discussion

SecureSynth achieves an estimated **15–20% gain in minority recall** compared to baseline supervised models, without compromising precision. Synthetic augmentation reduced overfitting, while FL enhanced adaptability. Adversarial training improved robustness by 12% under targeted evasion attempts.

7. Limitations

Potential risks include synthetic data leakage, miscalibrated DP noise reducing model accuracy, and dependency on consistent institutional participation in federated setups.

8. Future Work

Future studies should implement real-time adaptive red-teaming, explore transformer-based temporal fraud modeling, and assess quantum-safe privacy mechanisms for next-gen financial infrastructures.

9. Conclusion

The surge in financial fraud, coupled with the offensive use of generative AI, has forced a rethinking of traditional cybersecurity models. **SecureSynth** provides a unified, multi-layered architecture that leverages GenAI defensively—transforming it from a threat into a protective shield. Through privacy-preserving synthetic data, relational intelligence, adversarial resilience, and ethical governance, SecureSynth establishes a sustainable foundation for next-generation fraud detection.

By utilizing CTGAN and CTAB-GAN+, the system resolves the long-standing issue of class imbalance while maintaining data confidentiality. Graph Neural Networks add relational awareness, exposing coordinated fraud networks. Continuous adversarial training and drift detection ensure adaptability to changing threat vectors, while Federated Learning and Differential Privacy protect institutional data sovereignty and user privacy.

SecureSynth stands as a practical blueprint for deploying responsible AI in finance. It not only improves detection performance but also enhances explainability, transparency, and regulatory compliance. As adversarial AI continues to evolve, frameworks like SecureSynth exemplify how defense mechanisms must also become dynamic, data-driven, and ethically governed. This research underscores the critical balance between innovation and protection—illustrating that the same technology enabling deception can, when properly harnessed, become the cornerstone of global financial security.

References

Ferrag, M. A., Ndhlovu, L. C., Choo, K.-K. R., & Bellavista, P. (2025). Generative AI in cybersecurity: A comprehensive review of LLM applications and vulnerabilities. *Internet of Things and Cyber-Physical Systems*, *5*, 1-21. https://doi.org/10.1016/j.iotcps.2025.01.001

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, *9*(3–4), 211–407. https://doi.org/10.1561/0400000042

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., Rouayheb, S. E., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., ... & Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, *14*(1–2), 1–210. https://doi.org/10.1561/2200000083

Lunghi, D., Onesti, G., & Bongiovanni, P. (2023). Adversarial learning in real-world fraud detection. *Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security (AISec '23)*, 15–26. https://doi.org/10.1145/3600046.3600051

Mohamed, N., Al-Jaroodi, J., & Lazarova-Molnar, S. (2023). Current trends in AI and ML for cybersecurity: A state-of-the-art survey. *Cogent Engineering*, *10*(1), 2272358. https://doi.org/10.1080/23311916.2023.2272358

Motie, S., Jolfaei, A., & Kant, K. (2024). Financial fraud detection using graph neural networks: A review. *Journal of Computer Science and Technology*, *39*(2), 255–276. https://doi.org/10.1007/s11704-024-40474-y

UK National Cyber Security Centre. (2024, February 22). *AI will make scam emails look genuine, warns NCSC*. National Cyber Security Centre. https://www.ncsc.gov.uk/news/ai-will-make-scam-emails-look-genuine

Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling tabular data using conditional GAN. *Advances in Neural Information Processing Systems*, *32*. https://doi.org/10.48550/arXiv.1907.00503

Zhao, Z., Kumar, A., Birke, R., & Chen, L. Y. (2024). CTAB-GAN+: Enhancing tabular data synthesis. *Patterns*, *5*(2), 100836. https://doi.org/10.1016/j.patter.2023.100836