Sarcouncil Journal of Engineering and Computer Sciences

**ISSN(Online): 2945-3585**

Volume- 04| Issue- 11| 2025

**Research Article** **Received:** 05-10-2025| **Accepted:** 30-10-2025 | **Published:** 17-11-2025

# Fraud Detection in Banking Using Generative AI

***Dr. Sanjay Nakharu Prasad Kumar***

*IEEE Senior Member USA*

**Abstract:** Financial fraud continues to evolve in scale, sophistication, and speed, rendering traditional rule-based and supervised machine learning systems increasingly inadequate. This paper presents a comprehensive analysis of how generative artificial intelligence (AI) can transform fraud detection in banking by enabling proactive, adaptive, and highly scalable defense mechanisms. It examines the capabilities of key generative architectures—including Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), and transformer-based Large Language Models (LLMs)—and how they enhance anomaly detection, behavioral modeling, synthetic data generation, and unstructured text analysis. Real-world case studies from Swedbank, Mastercard, and JPMorgan demonstrate measurable improvements in detection rates, reduction of false positives, and faster identification of compromised accounts. The paper also discusses architectural considerations for deploying generative models at scale, addressing challenges related to adversarial attacks, explainability, privacy, and regulatory compliance. Finally, it explores emerging directions such as multimodal fraud detection, federated learning, adversarial defenses, and quantum-enhanced AI systems. By integrating generative AI with robust governance, scalable cloud architectures, and human oversight, banks can significantly strengthen their fraud detection capabilities and stay ahead of increasingly AI-enabled financial crime.

**Keywords:** Generative AI; Fraud Detection; Banking; Anomaly Detection; GANs; Variational Autoencoders (VAEs); Large Language Models (LLMs); Synthetic Data; Financial Crime; Anti-Money Laundering (AML); Cloud Architecture; RAG; Behavioral Modeling; Adversarial Attacks; AI Governance.

## INTRODUCTION

### Financial Fraud and Traditional Detection

An Introduction to Financial Fraud and Traditional Ways to Find It Identity theft, payment fraud, credit card abuse, and money laundering are all types of financial fraud that happen in banks. Institutions and customers lose billions of dollars each year because of these things (Deloitte Insights). The threat of landscape changes quickly. Scams that use generative AI, like deepfakes and fake identities, could make U.S. fraud losses go from $12.3 billion in 2023 to $40 billion by 2027 (Deloitte Insights). In the past, banks used supervised models and rules-based systems to find fraud. Rule engines use if-then heuristics or set thresholds to trigger alerts for transactions that are too big or come from countries that are on a blacklist (NVIDIA Technical Blog). These systems are easy to set up, but they have some clear problems.

Old-fashioned ways of finding things rely on known patterns and labeled examples. They frequently find it difficult to recognize new or changing schemes that do not conform to established fraud patterns (Dixit, A. 2024; Dixit, A. 2024). As fraud tactics get smarter, like smurfing funds in small amounts or using fake IDs, rule-based approaches don't work as well (NVIDIA Technical Blog). Every new scheme needs new rules, and the system might not be able to keep up with criminals' new ideas. Also, traditional methods have a lot of false positives, which annoy customers and raise the cost of investigations (Dixit, A. 2024; IBM). This lack of flexibility, along with more transactions, makes it harder for both human teams and static algorithms to keep an eye on things in real time (Financial Services Industry; IBM).

These problems show how important it is to have smarter and more adaptable fraud detection systems. Generative AI looks like a good way to fill in these gaps. It can learn complicated patterns and find things that fixed rules can't (Dixit, A. 2024; Dixit, A. 2024). Modern cloud architecture is also scalable, which means that AI-driven decision systems can now handle large numbers of transactions in real time (Kumar, S. N. P. 2025). The next sections talk about how generative AI models can help banks find fraud, how these systems are set up technically, how they are better than older methods, and the problems and case studies that come with them.

### Overview of Generative AI and Key Capabilities

Generative AI is a type of artificial intelligence that can make new, original content or data that looks like the patterns in the data it was trained on (IBM). Generative models learn the underlying distribution of input data and can create realistic samples of text, images, or transaction data. This is different from discriminative models, which only make predictions or classifications. Variational autoencoders (VAEs), generative adversarial networks (GANs), and transformer-based large language models (LLMs) are some of the most





***Corresponding Author:** Dr. Sanjay Nakharu Prasad Kumar

DOI- https://doi.org/10.5281/zenodo.17634095

important types of generative AI architectures (IBM).

**Variational Autoencoders (VAEs)**

VAEs are made up of a pair of networks called an encoder and a decoder that compress data into a latent representation and then rebuild it. By sampling from the learned latent space, they can make new outputs. VAEs have been employed for anomaly detection by acquiring the ability to reconstruct "normal" data; deviations (reconstruction errors) may signify fraud or outliers (Tang, T. *et al.,* 2025). VAEs offer a probabilistic framework to model authentic transaction patterns and identify those that deviate from them. Studies have shown that deep generative models, such as VAEs, are very good at finding strange things in complicated financial transactions (Tang, T. *et al.,* 2025). Advanced autoencoder architectures have demonstrated significant efficacy in enhancing fraud detection accuracy within credit card transactions (Kumar, S. N. P. 2025).

**Generative Adversarial Networks (GANs)**

In a minimax game, GANs use two neural networks: a generator and a discriminator (Dixit, A. 2024; Dixit, A. 2024). The generator makes fake data, like fake transactions, that tries to look like real data. The discriminator checks to see if the inputs are real or made up. Through this adversarial training, GANs learn to make samples that look very real, and at the same time, the discriminator gets better at finding strange things (Dixit, A. 2024).

GANs can model complicated, high-dimensional distributions of normal behavior and point out events that are outside of the normal range. A GAN can be trained on real transaction flows to find fraud. The discriminator then looks for transactions that are very different from the norm and flags them as possibly fraudulent (NVIDIA Technical Blog; Dixit, A. 2024). Conditional GANs and time-series GANs are two examples of GAN variants that can make fake fraud examples to add to sparse training data. It is important to note that GANs are sensitive to fraud patterns that are not well represented. After being trained on both real and GAN-simulated fraud cases, models are better at finding unusual transactions than models that were only trained on real data (Generative AI in Banking). Generative approaches have a big advantage because they can learn from data that isn't balanced, like when there are a lot more real transactions than fraud (Dixit, A. 2024).

**Large Language Models (LLMs)**

GPT-3 and GPT-4 are examples of modern LLMs that use transformer architectures. They are also generative, which means they mostly make text. LLMs are known for chatbots and creating content, but they can also look at unstructured data like emails, messages, and logs. They can even make summaries or explanations of fraud cases. For instance, JPMorgan created a system based on LLMs to find signs of email compromise fraud in internal communications (Deloitte Insights).

LLMs can help human fraud analysts by automatically scanning large documents or transaction narratives and pointing out suspicious patterns in plain English. The combination of Retrieval-Augmented Generation (RAG) architectures with LLMs has made them even better at giving contextually relevant fraud detection insights (Prasad Kumar, S. N. *et al.,* 2025; Kumar, S. N. P. 2025). Recent developments in cloud-optimized RAG architecture have made it easier to process large amounts of financial data more quickly (Kumar, S. N. P. 2025). Also, using advanced attention mechanisms to do sentiment analysis on text reviews has shown promise in finding fraudulent patterns in customer communications (Prasad Kumar, S. N. *et al.,* 2025).

As part of red-team exercises, LLMs can also make fake phishing messages or fraud scenarios to help train and stress-test fraud defense systems. LLMs basically take generative AI's reach beyond just numbers and into language, which is becoming more important in fraud that uses social engineering and narrative patterns.

**Key Capabilities Across Model Types**

Generative AI has important capabilities across different model types, including:

- **Unsupervised Learning:** Learning from large unlabeled datasets, such as millions of transactions, to understand what "normal" behavior looks like.
- **Anomaly Detection:** Identifying novelties or anomalies by assessing how well new observations match the learned model. Transactions that fit poorly within the learned distribution may suggest fraud.
- **Synthetic Data Generation:** Producing realistic but artificial data that retains the statistical properties of real data.





Publisher: SARC Publisher

The latter is useful for improving training by addressing data scarcity or class imbalance. It also helps with privacy. Banks can create synthetic customer data that reflects real patterns without revealing actual personal information. This allows for safer data sharing and model training while following privacy regulations (Generative AI in Banking). In summary, generative AI gives banks tools to identify and understand fraudulent behavior that they have not seen explicitly before. This represents a significant change from traditional rule-based systems.

**Applying Generative Models to Fraud Detection in Banking**

Generative AI models are utilized in various capacities to detect fraudulent activities in banking transactions and customer behavior. Principal applications encompass:

**Detection of Anomalies in Transactions**

Generative models such as GANs and VAEs can be utilized to identify anomalous transaction patterns in an unsupervised or semi-supervised approach. The objective is to model the distribution of legitimate transactions and subsequently identify outliers. A bank can train a GAN using its historical normal transaction data, such as daily payment flows. The GAN's discriminator functions as an anomaly detector, generating an anomaly score for each new transaction based on its likelihood according to the established "normal" profile (NVIDIA Technical Blog). Transactions exhibiting elevated anomaly scores are designated as suspicious for examination (NVIDIA Technical Blog).

Swedbank, one of Sweden's largest financial institutions, employed this methodology by training Generative Adversarial Networks (GANs) on an extensive dataset of 40 terabytes of transactions to identify patterns of money laundering and fraud. The GAN identified the patterns of legitimate transactions and could promptly notify on anomalous transfers in near real-time. GAN-based anomaly detectors proficiently discern intricate temporal and network patterns (e.g., sequences of fund transfers or clusters of associated accounts) that may signify fraud rings or laundering networks, which inflexible rules may overlook (NVIDIA Technical Blog).

Variational Autoencoders (VAEs) have been employed to reconstruct transaction features and assess reconstruction errors, enabling the detection of nuanced anomalies in customer spending patterns indicative of account takeover or misuse (Kumar, S. N. P. 2025). Research has integrated GANs and VAEs for this objective—Tang *et al.*, (2025) introduced a hybrid GAN-VAE framework in which the GAN produces credible normal transactions while the VAE guarantees that the latent space accurately reflects real data distributions, thereby enhancing the precision in detecting anomalous payments (Tang, T. *et al.*, 2025). Deep generative methods demonstrate markedly superior recall of infrequent fraud occurrences relative to traditional machine learning, particularly within extensive payment systems (Tang, T. *et al.*, 2025).

**Synthetic Fraud Generation and Data Augmentation**

Generative AI can help overcome the chronic problem of limited fraud examples for model training. Because fraudulent transactions are only a tiny fraction of all data, supervised classifiers often suffer from class imbalance. GANs offer a solution by simulating new fraudulent samples to balance the training set(Generative AI in Banking). For example, researchers have used GANs on credit card datasets to generate synthetic fraud transactions that resemble real fraud patterns (Generative AI in Banking).

By augmenting the training data with these GAN-generated instances, fraud detection models became more sensitive to underrepresented fraud behaviors and achieved higher detection rates than models trained on the original data alone (Generative AI in Banking). In one case, using a GAN-enhanced dataset improved fraud classification accuracy beyond what was achieved with the augmented dataset (Generative AI in Banking). Generative models can thus expose the classifier to a wider range of fraud scenarios, including hypothetical attacks that haven't yet occurred but are plausible.

Furthermore, generating synthetic data is useful for testing and validating fraud detection systems. Banks can create extensive what-if scenarios (e.g., coordinated fraud bursts, insider fraud cases) and ensure their detection pipeline flags them, all without risking sensitive customer data. Time-series generative models (like TimeGAN) can produce realistic temporal sequences of transactions, which are valuable for simulating long-term fraud behaviors or money laundering schemes for scenario analysis.



**Publisher: SARC Publisher**

**Behavioral Modeling and Customer Profiling**

Fraud often manifests as anomalies in a customer's behavior profile (sudden spending spree, atypical login pattern, etc.). Generative models can capture each customer's normal behavior signature and detect deviations. A VAE, for instance, could be trained per user (or per segment of users) on their transaction history to establish a personalized model of "normal" behavior, automatically flagging transactions that don't fit that user's profile.

GANs can similarly learn distributions of behavior across segments (e.g., typical daily spending patterns for salaried customers vs. retirees). Because generative models grasp joint patterns of multiple features (amount, time, merchant, location, device, etc.), they excel at spotting combinations of factors that look incongruent. This multivariate anomaly detection is crucial—a transaction that is individually reasonable (amount, location, time each within norms) might still be flagged because the joint pattern is unlike anything seen for that user (e.g., an unusual combination of merchant category and foreign location).

**Leveraging LLMs for Fraud Insights**

Large language models are being applied in fraud detection beyond text analysis; they can serve as AI assistants for fraud analysts. An LLM fine-tuned on fraud investigation reports and regulations can ingest an alert (for example, an anomalous account activity) and generate an explanatory report or next-step recommendations for investigators. This use of generative AI helps bridge the gap between raw data signals and human decision-making.

Additionally, banks use LLM-powered chatbots in customer-facing fraud prevention: for instance, a generative AI chatbot can engage with customers in real time when suspicious activity is noticed, asking adaptive verification questions or explaining why a certain transaction was flagged. The LLM's natural language generation capability enables a more conversational, context-aware fraud verification process, improving customer experience while security checks are performed.

Another novel application is using generative AI to produce honeypot content, e.g., fake phishing emails or fake dark web posts—to lure and identify fraudsters, or to train employees to recognize scams. The integration of advanced attention mechanisms, such as Random Multi-Hierarchical Attention Networks (RMHAN), has enhanced the ability to analyze sentiment and detect fraudulent patterns in textual data (Prasad Kumar, S. N. *et al.*, 2025).

**Technical Implementation of Generative AI Fraud Detection**

Deploying generative AI for fraud detection in banking requires careful consideration of model architecture, data handling, training procedures, and real-time deployment constraints. This section discusses the technical underpinnings:

**Model Architectures**

Each generative model used for fraud detection has a distinct architecture:

**GAN Architecture**: A typical GAN for transaction fraud consists of a generator G that takes random noise (and possibly conditional inputs like transaction context) and outputs synthetic transaction data, and a discriminator D that tries to distinguish real transactions from G's outputs (Dixit, A. 2024). Both G and Dare multi-layer deep neural networks (often fully connected or convolutional layers for tabular/sequence data). In training, D is optimized to correctly classify real vs. fake transactions, while G is optimized to fool D. Over time, G learns to produce highly realistic transaction patterns and D becomes adept at spotting subtle irregularities (Dixit, A. 2024).

For fraud detection usage, we typically retain the trained discriminator as the anomaly detector (since it encapsulates knowledge of what constitutes a normal vs. abnormal transaction) (NVIDIA Technical Blog). Some architectures integrate an encoder or use an autoencoder-GAN hybrid for stability and feature learning (NVIDIA Technical Blog). For example, the NVIDIA/Swedbank implementation based one of its GAN models on an unsupervised anomaly detection architecture that combined an encoder with GAN training to handle noisy labels and improve training stability (NVIDIA Technical Blog).

**VAE Architecture**: A VAE for fraud begins with an encoder network that compresses input transaction data x into a latent vector z (usually by outputting parameters $\mu(x)$, $\sigma(x)$ defining a probability distribution for z). A decoder network then samples z and reconstructs a transaction $\tilde{x}$. The training objective balances reconstruction accuracy with a regularization term pushing z to follow a standard normal distribution.





**Publisher: SARC Publisher**

After training on legitimate transactions, the VAE's decoder Dec(z) essentially represents the distribution of normal data. At deployment, for each new transaction x_new, the model computes its likelihood or reconstruction error. If x_new cannot be well reconstructed (i.e., falls in a low-density region of the learned latent space), it is labeled as anomalous. VAEs are relatively lightweight and can be deployed to process events in real time; their probabilistic nature provides a score of how normal or abnormal a transaction is (Kumar, S. N. P. 2025).

**LLM and Hybrid Architectures**: When using LLMs (transformers) for fraud analysis, the architecture might involve combining structured data and unstructured data. For example, a transformer encoder can process sequences of transactions (as a time series or as a token sequence after discretization of amounts, locations, etc.), possibly supplemented by textual metadata (like transaction descriptions or customer annotations).

LLMs pre-trained on general data can be fine-tuned with domain-specific corpora such as financial fraud case descriptions, suspicious activity reports, or regulatory guidelines. Cloud-optimized architectures enable efficient deployment of these large-scale models across distributed systems (Kumar, S. N. P. 2025). The integration of RAG architectures with LLMs has proven particularly effective in enhancing contextual understanding for fraud detection applications (Prasad Kumar, S. N. *et al.*, 2025; Kumar, S. N. P. 2025). Recent innovations in quantum-enhanced AI decision systems also show promise for future cloud-based machine learning applications in fraud detection (Kumar, S. N. P. 2025).

**Data Handling and Feature Engineering**

Financial transaction data is often high-dimensional and heterogeneous (timestamps, amounts, merchant codes, customer IDs, geolocation, device fingerprints, etc.). Before feeding into generative models, significant data preprocessing is needed. Common steps include data normalization or embedding (e.g., converting categorical features like merchant category or country into embeddings or one-hot vectors), feature engineering (creating aggregate features such as count of transactions in last 24 hours, or graph-based features capturing network connectivity between accounts (NVIDIA Technical Blog; NVIDIA Technical Blog)), and handling temporal dependencies (windowing sequences for recurrent models).

In the Swedbank GAN implementation, the bank used a feature store (Hopsworks) to engineer a rich set of features at scale (40 TB of data), including graph features that map relationships between entities (accounts, merchants) to detect complex fraud patterns (NVIDIA Technical Blog; NVIDIA Technical Blog). Graph features are powerful for uncovering rings or collusion (e.g., multiple individuals funneling money to a central account). The generative models can incorporate such features, for example by generating graph embeddings that correspond to realistic transaction networks and spotting anomalous subgraph patterns.

Cloud-based data engineering approaches have proven essential for handling the scale and complexity of modern financial datasets (Kumar, S. N. P. 2025). Advanced convolutional neural network architectures optimized through evolutionary algorithms have also shown effectiveness in processing and classifying complex transactional patterns (Preetham, A. *et al.*, 2024).

**Training Process**

Training generative models for fraud detection typically occurs offline on historical data due to the need for large datasets and intensive computation. Banks often leverage GPU clusters or cloud ML platforms for this task (NVIDIA Technical Blog). For example, training a GAN on millions of transactions may involve many epochs and careful hyperparameter tuning to ensure convergence (GANs are notorious for training instability like mode collapse).

Techniques such as Wasserstein loss, gradient penalty, or using ensemble discriminators can improve GAN training results for financial data. The Nvidia/Swedbank case reported nearly linear scaling of training throughput by using multiple GPUs in parallel, enabling them to train on tens of terabytes of data efficiently (NVIDIA Technical Blog) Scalable cloud architecture has become essential for supporting these computationally intensive training processes (Kumar, S. N. P. 2025).

An important aspect is validation: since unsupervised models don't optimize an obvious metric like classification accuracy, banks use proxy metrics (e.g., reconstruction error distribution, or how well known past frauds are



Publisher: SARC Publisher

assigned high anomaly scores by the model). They may also run simulated attacks through the trained model to evaluate detection performance.

**Real-Time Deployment Considerations**

Once trained, generative AI models must be deployed into the live transaction processing environment. Latency and throughput are primary concerns—the model should flag fraudulent transactions before they are completed or soon after, to allow intervention. Banks handle thousands of transactions per second; therefore, the fraud detection pipeline (including any generative model inference) needs to operate within a few milliseconds per transaction (Dixit, A. 2024).

Techniques to achieve this include model optimization (e.g., distilling a large model into a smaller one for inference, using lower precision arithmetic on GPUs or TPUs, or compiling models to efficient runtime engines) and scalable serving infrastructure (replicating the model across servers to handle load). The research by Dixit (2024) highlighted designing generative models to operate within milliseconds per inference so suspicious transactions can be halted immediately without bottlenecking legitimate flow (Dixit, A. 2024).

Cloud-optimized architectures for AI-driven decision systems have enabled real-time processing at scale (Kumar, S. N. P. 2025). Often, an event-stream processing framework is used: as transactions stream in, they are enriched with features, scored by the generative model, and if the anomaly score exceeds a threshold, an alert or block is triggered in real-time.

When cloud deployment is used, latency and data security must be carefully managed—many institutions choose on-premises or private cloud deployments for customer transaction data due to privacy and regulatory reasons (Dixit, A. 2024). Regardless of location, robust monitoring is required: drift detection (to see if model performance degrades as fraud patterns shift), uptime monitoring, and a fallback system (if the AI model fails or is offline, rule-based checks might temporarily take over to ensure continuity).

**Comparative Advantages of Conventional Approaches**

Generative AI-driven fraud detection offers several compelling advantages compared to conventional rule-based or discriminative machine learning methods:

**Detection of Novel Fraud Patterns**

Perhaps the greatest advantage is the ability to catch previously unseen fraud schemes. Traditional systems depend on known fraud signatures or human-crafted rules, making them largely reactive. In contrast, generative models operate by learning the normal data distribution and identifying anomalies without explicit prior examples (Dixit, A. 2024). This means they can flag suspicious behavior even if it does not match any known fraud pattern.

For example, a money laundering technique that involves a complex web of transfers might be recognized as anomalous by a GAN discriminator because it diverges from any patterns in legitimate transactions, despite not matching any rule in the database. In effect, generative AI provides a more proactive defense, crucial in an era when fraud tactics evolve rapidly. As noted in one study, a GAN-based framework was able to generalize across a wide range of financial behaviors and adapt dynamically to new fraud tactics, outperforming static models in recognizing emerging threats (Dixit, A. 2024).

**Reduced False Positives (Improved Precision)**

Generative models enable more nuanced pattern recognition, which can substantially lower false alarms. Rules are crude filters that often cast too wide a net (leading to many false positives) or too narrow (missing fraud). In contrast, AI models can consider myriad factors simultaneously and learn decision boundaries that better separate legitimate and fraudulent behavior.

The adversarial training process in GANs, for instance, refines the discriminator to be highly discerning, so that normal variability in customer behavior is not mistakenly flagged. Dixit (2024) reported that their GAN-based system significantly reduced the incidence of false positives compared to a legacy rules system (Dixit, A. 2024). By precisely identifying only truly suspicious activities, generative AI minimizes unnecessary customer disruptions.

As a concrete example, Mastercard's generative AI deployment doubled the detection rate of compromised cards while also reducing false positives by up to 200% (i.e., false positive alerts dropped to one-third of previous levels) in detecting fraudulent transactions on those cards (Dixit, A. 2024). This means banks can intervene faster with compromised accounts without inundating themselves (or their customers) with





Publisher: SARC Publisher

false fraud alerts. The improved precision not only cuts operational costs (fewer manual reviews) but also enhances customer satisfaction by avoiding wrongful transaction declines.

**Enhanced Coverage of Edge Cases and Imbalanced Data**

In fraud datasets, genuine transactions outnumber fraud by orders of magnitude, and certain fraud subtypes are extremely rare. Generative AI naturally addresses this because it doesn't require equal class representation. Models like GANs and VAEs pay attention to the entire distribution of data and can amplify the signal of rare events.

The GAN's generator can simulate underrepresented types of fraud to ensure the system learns them (Generative AI in Banking), and the discriminator can become sensitive to subtle cues from these cases. Research has shown that GAN-augmented models maintain high detection capability even when fraudulent patterns are sparse or previously unseen (Dixit, A. 2024). Advanced autoencoder architectures have demonstrated effectiveness in handling imbalanced fraud datasets (Kumar, S. N. P. 2025).

**Adaptability and Continuous Learning**

Once deployed, generative AI models can continue to improve by ingesting new data form of continuous or online learning. Unlike static rules that must be manually updated, AI models can be retrained or even updated in streaming fashion to adapt to fraudsters' evolving strategies.

Generative models are particularly adaptable: for instance, if fraudsters start changing their behavior to evade detection (an adversarial drift), an anomaly-detection GAN will immediately reflect that change in what it considers "normal" and thus still flag the new behavior as anomalous until it truly becomes mainstream (which gives banks a window to react). Additionally, the concept of adversarial training (pitting generator vs. discriminator) is effectively a constant learning mechanism (Dixit, A. 2024).

**Scalability to Big Data in Real Time**

AI-based systems scale through automation in a way human-centric processes cannot. Generative models, once trained, can score vast numbers of transactions quickly using parallel computation. Banks dealing with millions of transactions per day have found AI systems capable of real-time analysis across huge volumes that no manual team could handle (IBM)

The Swedbank GAN solution exemplified this by handling very large datasets with near-linear scaling on GPU clusters (NVIDIA Technical Blog; NVIDIA Technical Blog). Moreover, advanced generative models can be distributed across multiple servers or nodes—research has highlighted deploying GAN-based detectors in parallel across distributed systems to ensure robust, speedy fraud detection without compromising accuracy (Dixit, A. 2024; Dixit, A. 2024). Scalable cloud architecture has proven essential for supporting these large-scale deployments (Kumar, S. N. P. 2025).

**Improved Fraud Loss Savings**

Ultimately, the combination of higher detection rates and lower false positives yields tangible financial benefits. Faster detection prevents more fraudulent transactions from completing, directly reducing losses. For instance, AI fraud systems have helped large financial institutions save significant sums—one global bank reportedly saved $150 million in a single year after deploying AI fraud detection techniques (NVIDIA Technical Blog; NVIDIA Technical Blog).

## CHALLENGES AND RISKS

Despite its promise, the use of generative AI in fraud detection comes with challenges that banks must carefully manage. Key issues include adversarial risks, model explainability, data privacy, and regulatory compliance:

**Adversarial Risks and Criminal Use of AI**

The rise of generative AI is a double-edged sword—while banks use it to detect fraud, criminals can use it to perpetrate fraud. Adversaries may exploit AI systems through adversarial attacks. For instance, fraudsters can probe a bank's AI model by submitting transactions with slight modifications designed to evade detection, a practice known as adversarial evasion. They might also attempt to "poison" the training data, injecting deceptive records so the model learns incorrect patterns (Financial Services Industry).

Furthermore, generative AI itself provides tools for criminals: cheaply produced deepfake videos or voice clones have enabled social engineering heists (e.g., AI-synthesized voices of executives tricking employees to transfer funds) (Generative AI in Banking). The availability of generative models for creating synthetic identities, complete with realistic fake documents and credit histories, has lowered the barrier for identity fraud and loan



Publisher: SARC Publisher

application fraud (Financial Services Industry). Phishing campaigns can now be automated at scale with AI-generated, personalized emails and even phone calls (voice deepfakes), making them more convincing and harder to filter (Financial Services Industry).

All these developments demand that banks not only fortify their models against direct attacks but also broaden their defenses to counter AI-augmented fraud attempts. It becomes essential to continuously update detection logic (possibly using adversarial training techniques to anticipate how criminals might try to fool the AI (Dixit, A. 2024)) and to incorporate multi-factor checks that are harder for AI to spoof.

**Explainability and Model Transparency**

Generative models, especially deep neural networks, are often criticized as "black boxes." In a highly regulated domain like banking, explainability is crucial—banks must be able to explain why a transaction was flagged or a customer was denied service, both for customer communication and for regulator audits. Traditional rules have the benefit of transparency, whereas a deep GAN or VAE might flag an anomaly without an obvious human-interpretable reason.

To address this, researchers and practitioners are incorporating explainability techniques into AI fraud systems. One approach is feature attribution, which analyzes the trained model to identify which input features most influenced a particular alert (Dixit, A. 2024). For example, if a transaction is flagged by the model, an attribution method (like SHAP values) might reveal that an unusual device ID and a large transaction amount combined were the top factors.

The framework introduced by Dixit (2024) explicitly integrated interpretability mechanisms so that the decision-making process of the GAN-based model can be explained to regulators on demand (Dixit, A. 2024). This transparency is critical not only for regulatory reasons but also to ensure bias has not crept into the model.

**Data Privacy and Security**

Using customer data to train generative models raises significant privacy concerns. Transaction and account data are highly sensitive, and regulators (as well as customers) demand that privacy be preserved. A major challenge is that training large AI models often requires centralized data aggregation, which could conflict with data residency laws or internal policies.

Banks are exploring privacy-preserving techniques to mitigate this. One strategy is using synthetic data: generative models themselves can generate artificial datasets that mirror real data's statistical properties without revealing individual personal data (Generative AI in Banking)(Generative AI in Banking). This synthetic data can be used for model training, development, or sharing with external vendors without risking PII exposure.

Another strategy is federated learning or on-device learning, where the model training happens locally at each data source and only aggregated updates (not raw data) are sent to a central server. Additionally, banks are adopting advanced encryption techniques like homomorphic encryption when deploying AI in the cloud—this allows computations on encrypted data so that cloud servers never see raw transaction details (Dixit, A. 2024).

Cloud-based architectures with robust security measures have become essential for protecting sensitive financial data while enabling AI-driven decision systems (Kumar, S. N. P. 2025)(Kumar, S. N. P. 2025).

**Regulatory and Ethical Concerns**

Financial regulators are keenly aware of AI's growing role and are increasingly scrutinizing its use in fraud prevention. Key regulatory concerns include model fairness, accountability, and alignment with existing laws (like AML regulations). For example, if a generative model flags suspicious transactions for Anti-Money Laundering, the bank must still file SARs (Suspicious Activity Reports) that regulators can understand.

In the United States, the OCC and Federal Reserve have issued guidance on model risk management for AI, implying banks should have processes to validate models and prevent uncontrolled use of AI in decisions that affect customers. Compliance with financial regulations is non-negotiable—any AI-based fraud system must still achieve the outcomes regulators expect.

The advantage is that if generative AI improves precision, it can help banks exceed regulatory requirements (catching more illicit activity and reducing false flags that burden investigative units)(Dixit, A. 2024). Indeed, Dixit (2024) notes that the GAN framework enhances precision and





Publisher: SARC Publisher

robustness in line with regulations to prevent money laundering and terrorist financing (Dixit, A. 2024).

Regulators also demand documentation: banks should document how the model was trained, what data was used, and how it has been validated—essentially treating the AI like any other critical risk model. Transparency to regulators can be improved by providing them with simplified descriptions or by using AI audit tools that output compliance reports.

**Case Studies and Recent Implementations**

Leading banks and payment companies have begun implementing generative AI techniques in their fraud detection workflows, with notable success. Below we highlight a few real-world case studies and implementations:

**Swedbank – GANs for Anomaly Detection**

Swedbank, one of the largest banks in Sweden, developed a cutting-edge fraud detection solution using GANs as a core component (NVIDIA Technical Blog). Faced with massive data volumes and complex money-laundering schemes, Swedbank collaborated with tech partners to train GAN models on a 40-terabyte dataset of transactions (NVIDIA Technical Blog).

Their approach treated fraud detection as a semi-supervised anomaly detection problem: the GAN's generator learned the patterns of normal (legal) transactions, and the discriminator was used to spot abnormal transactions that didn't fit those patterns . They also incorporated graph analytics: transactions were modeled as a graph of entities (individuals and businesses with edges representing fund flows) to catch structures typical of laundering (e.g., "gather-scatter" patterns where funds concentrate then redistribute) (NVIDIA Technical Blog; NVIDIA Technical Blog).

By leveraging NVIDIA GPUs and the Hopsworks platform, Swedbank was able to train and deploy these deep models at scale, achieving near real-time detection despite the data size (NVIDIA Technical Blog). This GAN-based system enabled the bank to identify complex fraud patterns and trigger alerts much faster than previous methods. While exact performance metrics are confidential, it was reported that such AI-driven approaches contribute to substantial fraud loss reductions—aligning with industry reports of large banks saving tens of millions annually via AI fraud prevention (NVIDIA Technical Blog).

**Mastercard – Generative AI for Compromised Card Detection**

Mastercard, a global payments company, deployed generative AI technology to accelerate credit card fraud detection across its network (Mastercard accelerates card fraud detection with generative AI technology; Dixit, A. 2024). One of Mastercard's challenges is identifying when card details have been compromised (e.g., via breaches or skimming) so that issuers can be alerted to block those cards before fraud occurs.

In 2024, Mastercard announced a new generative AI-based predictive system that scans billions of card transactions and millions of merchants to find patterns indicating a card may be compromised (Mastercard accelerates card fraud detection with generative AI technology;Dixit, A. 2024). According to Mastercard, this system doubled the speed of identifying compromised cards and enabled blocking them much sooner than previous methods (Mastercard accelerates card fraud detection with generative AI technology). In addition, it reduced false positive alerts for fraud by up to 200% and made merchant risk identification three times faster (Dixit, A. 2024).

In practice, this means when fraudsters steal partial card data and attempt to use or sell it, Mastercard's AI can piece together clues from transaction streams to predict the full card number or at least flag the card as likely compromised, prompting proactive cancellation. The results are impressive: faster response (reducing the window in which fraudsters can misuse the cards) and higher accuracy (so cardholders are less often impacted by erroneous fraud blocks).

**JPMorgan Chase – LLMs for Scam Detection**

JPMorgan Chase, the largest bank in the US, has been experimenting with large language models to combat fraud in areas like email compromise and phishing. Deloitte reported that some banks (including JPMorgan) are incorporating LLMs to detect signs of fraud in communications, such as using an AI to scan internal emails for social engineering attempts (Generative AI in Banking).

One specific application is detecting Business Email Compromise (BEC) scams, where a fraudster impersonates a company executive via email to trick employees into wiring money. JPMorgan's use of an LLM likely involves analyzing email language and context to catch anomalies—e.g., an email that looks like it's from





Publisher: SARC Publisher

the CFO but whose wording or timing is inconsistent with the CFO's normal behavior.

An LLM can be trained in both legitimate communications and known scam messages to generate a risk score for incoming emails. This adoption shows that fraud detection isn't limited to transactional data—customer and employee communication channels are also protected using AI. The integration of advanced RAG architectures and attention mechanisms has enhanced the capability to analyze and understand complex communication patterns (Prasad Kumar, S. N. *et al.,* 2025; Kumar, S. N. P. 2025)

**Research Prototypes and Emerging Solutions**

Beyond these production deployments, numerous prototypes and academic/industry collaborations are pushing generative AI in finance. For example, researchers developed a "Lightweight GAN" model for real-time credit card fraud detection that operates on edge devices with limited computers, making AI fraud screening feasible at the point-of-sale or ATM level (Advanced R-GAN; arXiv preprint)

Another notable direction is Variational Graph Autoencoders for fraud—combining graph neural networks with generative models to detect fraud rings. Companies like Feedzai and FeatureSpace, which provide fraud AI solutions, are integrating generative components (like synthetic data generation modules) to improve their systems' adaptability.

Advanced autoencoder architectures combined with deep neural networks have shown promise in improving fraud detection accuracy (Kumar, S. N. P. 2025). Optimized convolutional neural network approaches using evolutionary algorithms have also demonstrated effectiveness in pattern recognition tasks relevant to fraud detection (Preetham, A. *et al.,* 2024).

**Conclusion and Future Directions**

Generative AI is poised to redefine fraud detection in banking by combining the analytic power of machine learning with creative simulation capabilities, enabling banks to stay ahead of increasingly cunning fraudsters. The evidence so far—from research studies to industry deployments—indicates that generative AI can significantly enhance detection rates (especially for novel and sophisticated schemes) while reducing false positives (Tang, T. *et al.,* 2025; Mastercard accelerates card fraud detection with generative AI technology). It offers greater adaptability, as models can learn and evolve with minimal human reprogramming, which is crucial in the ever-shifting fraud landscape (Dixit, A. 2024)

Banks leveraging these technologies have reported faster response times and substantial savings by preventing fraud losses and reducing manual review effort (Dixit, A. 2024; NVIDIA Technical Blog). Looking forward, several future directions are likely to shape this field.

**Integration of Multimodal AI**

Future fraud detection systems may combine generative models across different data modalities—linking transactional data with text (communications), images (e.g., IDs or checks), and even audio (call center recordings). For example, an advanced system could use a VAE to flag a suspicious transaction, an LLM to cross-read the customer's recent communications for context, and a generative image model to verify the authenticity of documents provided.

Such a multimodal AI approach would give a 360-degree analysis of fraud cases, improving accuracy. This requires further research in fusing outputs of different generative models and orchestrating them in real time. The integration of advanced attention mechanisms and RAG architecture shows promise for coordinating multiple AI models effectively (Prasad Kumar, S. N. *et al.,* 2025; Kumar, S. N. P. 2025)

**Explainable and Ethical AI by Design**

Given regulatory pressure, we anticipate a stronger emphasis on explainable AI frameworks for generative models. Future models might have built-in interpretability, for instance, new architectures or training methods that produce human-interpretable features (there is emerging research on "disentangled" VAEs that isolate meaningful factors in the data).

Additionally, banks will develop standardized ways to document AI decision logic for auditors. Ethical AI training (ensuring models don't pick up biases from historical data) will also be key, possibly through techniques like fairness constraints during model training. The goal is for generative AI to not only be powerful but also transparent, fair, and accountable, aligning with the concept of Responsible AI that many financial institutions are now championing.





**Publisher: SARC Publisher**

**Adversarial Defense and Model Robustness**

As adversaries evolve, so must the defenses. We foresee more use of adversarial training (training fraud models on adversarial examples or with simulated attacker strategies) to harden them. Research into GANs for cyber defense is already underway, using one AI to simulate attack patterns and another to learn to detect them (Dixit, A. 2024).

In fraud detection, this could mean continuously generating new fraud scenarios (via a generator) to challenge the detector. Moreover, ensemble approaches will be used to increase robustness: multiple diverse models (some generative, some not) can cross-verify decisions such that an attacker would need to evade all simultaneously, which is far harder.

Model monitoring tools will get smarter at spotting when an AI model's behavior changes (possibly due to concept drift or adversarial influence), triggering retraining or human review. In essence, an arms race is in progress, and future systems will likely have self-correcting abilities to maintain resilience against adversarial AI attacks.

**Federated and Privacy-Preserving AI Collaboration**

Fraud is often a cross-institution problem, the same fraudster might hit multiple banks. However, data privacy concerns make sharing raw data difficult. Future solutions may employ federated learning or secure multi-party computation to jointly train generative fraud models on combined datasets of many banks without sharing sensitive data.

For instance, a consortium of banks could train a global GAN where each bank's data stays local and only model parameter updates are shared and aggregated. This way, the model learns a wider variety of fraud patterns from across institutions, benefiting everyone, while each bank's data remains confidential. Techniques like homomorphic encryption and differential privacy will underpin these collaborations to ensure regulatory compliance when models span geographical and organizational boundaries (Dixit, A. 2024).

The result could be industry-wide AI fraud networks that detect coordinated attacks that no single bank could have caught in isolation. Cloud-based architectures with robust security and privacy controls will be essential for enabling such collaborative approaches (Kumar, S. N. P. 2025).

**Real-Time Adaptive Interventions**

As detection becomes faster with AI, the next step is automated or semi-automated response. Future generative AI systems might not only flag fraud but also take action—for example, generating an immediate challenge to the user ("We noticed an unusual transaction, please confirm X or provide additional authentication") or even generating a honey-token response to engage the fraudster.

Generative AI could personalize these interventions. For legitimate customers, it might generate a polite, context-aware explanation of why an action is needed (improving customer experience even during a security check). For suspected fraudsters, it might generate dynamic content to confuse or draw them out (an approach used in cybersecurity deception).

While full automation must be approached cautiously, it is an area of growth, moving from AI in the loop to AI in charge for certain fraud scenarios, under human oversight. Advanced sentiment analysis and natural language generation capabilities will enhance the effectiveness of these adaptive interventions (Prasad Kumar, S. N. *et al.*, 2025).

**Regulatory Frameworks and AI Governance**

In the future, we expect clearer regulatory frameworks specifically addressing AI in fraud and financial crime. This might include certification of AI models, regulatory tech interfaces for AI outputs (e.g., AI-generated SAR filings), and standardized testing procedures.

Banks will need to align their generative AI implementations with these frameworks, which may involve regular audits, stress-test the AI (akin to model stress tests), and demonstrating compliance through technical documentation. Generative AI systems might also be used by regulators themselves as supervisory technology—for example, regulators could deploy their own generative models to analyze industry data for systemic fraud risks or to validate the efficacy of banks' models (Generative AI in Banking).

Therefore, banks should prepare for a future where AI competency is not just a competitive advantage but a regulatory expectation. Investing in internal AI governance (boards, committees, and controls that oversee AI use) will be as important as the technology itself.





**Publisher: SARC Publisher**

**Quantum-Enhanced AI Systems**

Emerging quantum computing technologies show promise for enhancing AI-driven fraud detection systems. Quantum-enhanced architectures could potentially process complex fraud patterns more efficiently and handle the massive computational requirements of advanced generative models (Kumar, S. N. P. 2025). While still in early stages, research into quantum machine learning applications for financial services suggests significant potential for future fraud detection capabilities.

**Advanced Cloud-Optimized Architectures**

The continued evolution of cloud-based infrastructure will enable more sophisticated AI-driven decision systems. Scalable cloud architectures optimized for machine learning workloads will support real-time processing of massive transaction volumes while maintaining low latency (Kumar, S. N. P. 2025). Healthcare and other industries have already demonstrated how cloud-based AI systems can transform decision-making processes, providing valuable lessons for financial services (Kumar, S. N. P. 2025).

**Final Thoughts**

In conclusion, fraudsters will undoubtedly continue to leverage technology to find new weaknesses, but generative AI provides a formidable counterforce. The banking industry's embrace of generative models marks a shift toward more intelligent, adaptive, and data-driven defense mechanisms. The journey is ongoing—challenges around trust, ethics, and security remain—but the trajectory is clearly toward AI-enhanced fraud prevention that can safeguard the financial system in real-time.

By combining generative AI's capabilities with sound governance and human expertise, banks can significantly strengthen their fraud-fighting arsenal. The future likely holds a collaborative ecosystem where humans and AI work hand-in-hand: AI sifts oceans of data to pinpoint threats, and human experts provide strategic oversight and handle the complex cases or ethical judgments.

Such synergy will be essential to maintain trust in the financial system as both commerce and crime become increasingly digital and automated. Generative AI, used wisely, will help ensure that as the fraudsters get smarter, the defenders do too—staying one step ahead to protect customers and institutions from financial crime.

## REFERENCES


1. Deepfake banking and AI fraud risk. Deloitte Insights. (Deloitte Insights)
2. Detecting Financial Fraud Using GANs at Swedbank with Hopsworks and NVIDIA GPUs NVIDIA Technical Blog. https://developer.nvidia.com/blog/detecting-financial-fraud-using-gans-at-swedbank-with-hopsworks-and-gpus/
3. TechRxiv preprint, "Generative AI for Financial Fraud Detection." TechRxiv preprint. https://d197for5662m48.cloudfront.net/documents/publicationstatus/229119/preprint_pdf/96f1521d3912da331a7a651f0c5c983a.pdf
4. AI Fraud Detection in Banking | IBM. https://www.ibm.com/think/topics/ai-fraud-detection-in-banking
5. What is Generative AI? | IBM. https://www.ibm.com/think/topics/generative-ai
6. Tang, T., Yao, J., Wang, Y., Sha, Q., Feng, H., & Xu, Z. "Application of Deep Generative Models for Anomaly Detection in Complex Financial Transactions." *2025 4th International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID)*. IEEE, (2025).
7. Top 6 Use Cases of Generative AI in Banking. https://research.aimultiple.com/generative-ai-in-banking/
8. Deepfake banking and AI fraud risk | Deloitte Insights. https://www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html
9. Mastercard accelerates card fraud detection with generative AI technology. https://www.mastercard.com/us/en/news-and-trends/press/2024/may/mastercard-accelerates-card-fraud-detection-with-generative-ai-technology.html
10. How Cybercriminals Are Exploiting AI in the Financial Services Industry. https://www.citrincooperman.com/In-Focus-Resource-Center/How-Cybercriminals-Are-Exploiting-AI-in-the-Financial-Services-Industry
11. Advanced R-GAN: Generating anomaly data for improved detection. https://www.sciencedirect.com/science/article/pii/S1110016824012523
12. arXiv preprint. “Utilizing GANs for Fraud Detection: Model Training with Synthetic






**Publisher: SARC Publisher**

Data." *arXiv preprint*. https://arxiv.org/pdf/2402.09830
13. Kumar, S. N. P. "Scalable Cloud Architectures for AI-Driven Decision Systems." *Journal of Computer Science and Technology Studies* 7.8 (2025): 416-421.
14. Prasad Kumar, S. N., Gangurde, R., & Mohite, U. L. "RMHAN: Random Multi-Hierarchical Attention Network with RAG-LLM-Based Sentiment Analysis Using Text Reviews." *International Journal of Computational Intelligence and Applications* (2025): 2550007.
15. Kumar, S. N. P. "AI and Cloud Data Engineering Transforming Healthcare Decisions." *Journal Of Engineering And Computer Sciences* 4.8 (2025): 76-82.
16. Kumar, S. N. P. "Recent Innovations in Cloud-Optimized Retrieval-Augmented Generation Architectures for AI-Driven Decision Systems." Engineering Management Science Journal, Vol. 9, No. 4. (2025).
17. Kumar, S. N. P. "Quantum-Enhanced AI Decision Systems: Architectural Approaches for Cloud-Based Machine Learning Applications." SAR Council. (2025). https://sarcouncil.com/2025/08/quantum-enhanced-ai-decision-systems-architectural-approaches-for-cloud-based-machine-learning-applications
18. Preetham, A., Vyas, S., Kumar, M., & Kumar, S. N. P. "Optimized convolutional neural network for land cover classification via improved lion algorithm." *Transactions in GIS* 28.4 (2024): 769-789.
19. Kumar, S. N. P. "Improving Fraud Detection in Credit Card Transactions Using Autoencoders and Deep Neural Networks." *Diss. The George Washington University*, (2022).

**Source of support:** Nil; **Conflict of interest:** Nil.

**Cite this article as:**
Kumar, S. N. P. "Fraud Detection in Banking Using Generative AI." *Sarcouncil Journal of Engineering and Computer Sciences* 4.11 (2025): pp 133-145.





**Publisher: SARC Publisher**