International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| Volume 5, Issue 4, 62-73, 2024
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V5I4P107

Original Article

AI-Driven Security for Financial Transactions: Leveraging LLMs, Federated Learning, and Behavioral Biometrics

Anitha Mareedu
Electrical engineering Texas A&M university - Kingsville 700 University Blvd, Kingsville, TX 78363.

Abstract *- The rising sophistication of cyber threats ranging from phishing and synthetic identities to adversarial model attacks has made the demand for intelligent, adaptive security solutions in financial systems more urgent than ever. This study explores key AI technologies that are shaping the future of secure financial transactions, including Large Language Models (LLMs), Federated Learning (FL), Graph Neural Networks (GNNs), and behavioral biometrics. For each technology, we outline its core architecture, operational mechanisms, and applicability in real-world fraud detection systems. LLMs enable contextual understanding of transaction narratives, aiding in the detection of phishing attempts across various communication channels. FL facilitates collaborative model training across multiple financial institutions without compromising user privacy. GNNs leverage the relational structure of transaction networks to uncover fraud rings that evade traditional rule-based systems. Behavioral biometrics offers continuous authentication by analyzing passive user attributes such as typing patterns and device interaction. A comparative analysis demonstrates the advantages of these AI approaches over conventional methods, highlighting improvements in detection accuracy, scalability, and privacy preservation. The review also addresses critical challenges including data imbalance, latency, model drift, and regulatory constraints. Together, these insights provide a comprehensive foundation for understanding how AI, when applied responsibly, can enhance the integrity and resilience of financial ecosystems.*

Keywords *- large language models (LLMs), federated learning (FL), graph neural networks (GNNs), privacy-preserving AI, explainable AI (XAI), regulatory compliance, cybersecurity in finance.*

1. Introduction

The financial sector's digitalization has picked up speed in the last few years. With as many efficiencies and more access to financial services that these technologies brought, these technologies also subjected the industry into an increasing sophistication of cyber threats [1][2]. Financial frauds which started off as straightforward phishing and malware have become advanced to include synthetic identity frauds [3], adversarial machine learning, and deepfake-supported impersonations [4] . The new threats exploit the weaknesses of the Traditional security system, which tends to rely on fixed rules, fixed models of detection, and retrospective examination. The growing popularity of synthetic identity fraud may be, perhaps, the most subversive trend throughout this period where the offenders establish identities by combining both actual and false personal details. These identities are typically unidentified over long periods of time and this enables large-scale fraudulent activities at a later stage in time[7]. At the same time, robbers have polished their phishing tricks, thus being nearly official.

Now they rely on artificial intelligence texts, deep-fakes and straightforward social clues to find the past through security gates, betray trust and pick up the login credentials. The numbers of business email compromise (BEC), fraudulent wire transfer requests and fraudulent transaction approvals are on the rise, according to their volume and impact [5] . Consequently, financial institutions such as banks are adopting this kind of adaptive and smarter security due to the emergence of AI (artificial intelligence) and ML (machine learning) [6]. The shift in the salient direction to tools that incorporate behavior learning, real-time threat modification, and data processing adherent to the culture of privacy is already in progress. Three AI-powered strategies now stand out as game changers for safeguarding financial deals: large language models (LLMs), federated learning, and behavioral biometrics[8]. They have separate and distinct points to attack but through the collective efforts, better fraud identification, continuing authentication, and mutual sharing of intelligence between banks and vendors are established. The following table 1 gives a glimpse of this changing toolkit.

Table 1: Evolution of AI Techniques in Financial Security

Key Advancement	Description	Application Area
Rule-Based AI & Supervised ML [9]	Traditional fraud detection using labeled datasets and predefined rules	Transaction risk scoring
Federated Learning Adoption [10]	Collaborative learning across banks without centralizing user data	Privacy-preserving fraud detection

Behavioral Biometrics[11]	Use of user behavior traits for continuous, real-time authentication	Identity verification and account protection
Integration of LLM [12]	LLMs applied to detect phishing, analyze narratives, and enhance automation	Scam detection, natural language fraud analysis

Large language models such as the GPT-3 model developed by OpenAIs, and the BERT developed by Google, which were initially designed to process ordinary language, are being used to read and cull bank emails, chat messages, and transaction notes. Their ability to interpret context and linguistic patterns made them suitable for identifying phishing attempts, fraud indicators in messages, and anomalies in transaction descriptions. Federated learning enabled multiple institutions to collaboratively train models without exchanging sensitive data, thus balancing intelligence sharing with data privacy compliance, a key concern in regulated financial environments. At the same time, behavioral biometrics quietly watch how each user types, scrolls, and taps on a screen, creating an ongoing, behind-the-scenes ID check.

Together, these tools deliver swift, hard-to-fake fraud alerts that thieves find far easier to dodge in theory than in practice. This review focuses exclusively on a defined period of recent technological advancements, deliberately excluding emerging innovations that fall outside the scope of the analysis, such as generative agent frameworks and post-quantum cryptographic systems. By concentrating on technology that has already gone live, the analysis seeks to deliver a tight yet thorough appraisal of proven, AI-led tactics that banks and payment firms now deploy. It centers on core tools and surveys how they have measurably strengthened fraud defenses, secured individual transactions, and built wider public trust in online money systems.

1.1. Research Objective:

Here are the three research objectives:

- To investigate AI-driven approaches (LLMs, FL, GNNs, behavioral biometrics) for enhancing financial fraud detection from 2020 to 2023.
- To evaluate the performance, privacy, and effectiveness of these AI techniques compared to traditional security methods.
- To assess the ethical, regulatory, and implementation challenges associated with deploying AI in financial transaction security.

2. Threat Landscape in Financial Transactions

The financial services industry remains one of the most attractive targets for cybercriminals due to the high liquidity of assets and the sensitivity of transactional data [13]. A new threat added to the already cluttered threat landscape comes in the form of digital banking, smartphone wallets, and web-based investment tools. Malwares and denial-of-service attacks, which used to be old threats, now come equipped with the benefits of machine learning and AI and security personnel continue to update these vintage defense mechanisms. With banks and fintechs cramming more features into mobile apps, the bad guys develop more subtle, scalable, and highly targeted plans that sneak through minute software loopholes.

2.1. Traditional Attack Vectors

The use of traditional methods of financial frauds has also remained a big threat especially in systems that use simple methods of authentication. The use of card-not-present (CNP) fraud has persisted in online shopping and e-commerce activities and has in most cases dodged normal verification procedures [14] . Phishing attempts are continuing to cause successful selections of users after years of sensitization campaigns by presenting realistic spoof emails, pretend websites, and customer support scam. Such attacks are often used to steal credentials or initiate fraudulent funds transfer under the pretext of user trust. Another form of attack that is persistent in the current mobile-banking era is SIM-swapping. In such an arrangement, the criminal convinces a telecommunication expert to port the number of the victim to a SIM that is under their control. After that, the scammer receives one-time codes and two-factor prompts, and thus they get access to financial accounts. Likewise, ATO raids are still present and are usually perpetrated by credential-stuffing bots which rely on stolen credentials and reoccurring passwords.

2.2. Emerging Threats

Since the application of artificial intelligence and machine learning technologies to attack systems and defence systems, fraud has taken new and more sophisticated evasive forms. A remarkable change is the utilization of deepfake technology to pose as customers, executives and employees by using synthetic audio or video that doesn't look unreal. Such tools have been used to tamper with verification systems to engage in high value social engineering attacks on financial organizations [15]. An emerging issue in the fraud space is synthetic identity theft, whereby thieves cleverly combine actual and fabricated information to create completely fictional identities. After they have been created, such fake identities can easily pass through routine measures of verification and progressively accumulate a virtually legitimate background before initiating fraud. The

profiles made need not be based on a real person and are, therefore, not flagged by standard watch lists or background checks since they were newly made up and not modified based on someone.

Besides, adversarial machine learning has also resulted in a novel source of weaknesses. The attackers have even more power to slightly modify the input to trick the fraud detection programs since the inputs are now more subtle, e.g., metadata of transactions or behavioral tells. Such adversarial examples attack the vulnerabilities of the decision boundaries of machine learning models and result in the inference of the use of malicious behaviour as benign. This has brought forth deep concerns on the effectiveness and explainability of AI-founded security systems in the world of high stakes in the financial industry.

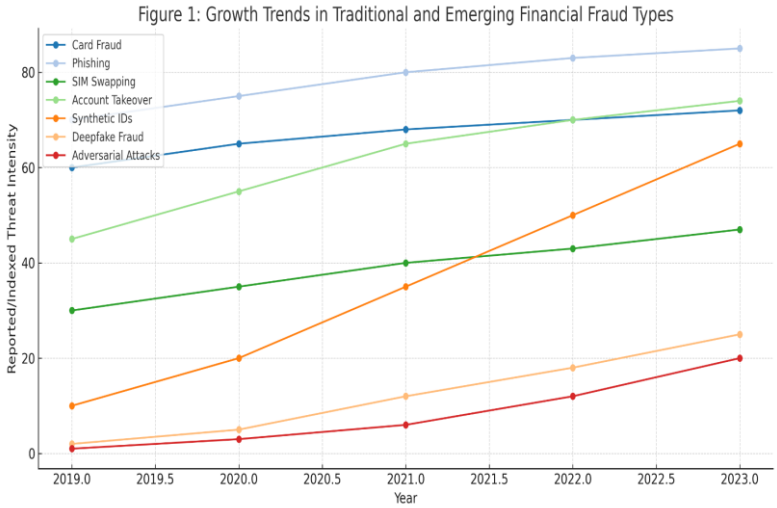

Figure 1. Growth trends in traditional and emerging financial fraud types.

Figure 1 shows the proportionate increase in the different categories of financial fraud where it can be seen that both the traditional and emerging threat vectors have shown a proportionate increase. It is important to note that the figure shows the increasing popularity of newer threats like synthetic identity fraud, AI-enhanced deception along with older threats like phishing, and card fraud still remaining most popular. This comparison depicts increasing attack surface and the requirement of adaptive defense mechanisms to deal with older and emerging fraud techniques. Today social engineering, behavioral baiting and algorithmic abuse are no longer functioning independently and thus the threat landscape is more intricate than ever. The right way to address that challenge would be to implement a mixed approach: real-time behavior monitoring, biometrics, and powerful AI models that can detect the smallest deviations, even when adversaries attempt to counter them.

3. Large Language Models (LLMs) for Financial Security

Large Language Models (LLMs) became an effective way of identifying and preventing language-related threats in financial systems. Their performance of reading and creating man-like text allows more sophisticated uses in transaction monitoring as well as fraud detection and threat intelligence [16].

3.1. Introduction to LLMs: GPT-2, GPT-3, T5, BERT, RoBERTa

The use of Large Language Models (LLMs) is also becoming essential in financial cyber security since these models are able to read, parse, and interpret dense natural-language text [18] . Even though such models have initially been designed to be applied in mundane pursuits such as machine translation, question-answering and summarising, practitioners are rapidly re-appropriating them to interpret user chats, transactions notes, and other digital interactions that require a deeper semantic understanding. Among the most widely adopted models are GPT-2, GPT-3, T5, BERT, and RoBERTa. These transformer-based models vary in their construction and training principles (GPT models are autoregressive whereas BERT-based models are bidirectional and it is designed to be trained to perform classification tasks). What they have in common is their contextual awareness that allows them to analyze unstructured data nuances that cannot be analyzed using the traditional rule-based or keyword-matching systems. In financial cybersecurity, LLMs are employed to detect anomalies in messages and transaction metadata, monitor communications for suspicious behavior, and flag potential phishing attempts across email, SMS, and voice transcripts. Their contextual reasoning capabilities are critical in identifying subtle linguistic manipulations that could signal fraud or unauthorized activity.

3.2. LLM Applications in Financial Security

Large language models (LLMs) can be re-tasked to do a variety of things, and today industry companies are relying on the technology in several areas of financial security [17]. Filtering through transfer notes, providing live assistance to frontline personnel, or detecting social engineering attacks, these prototypes incorporate intelligent context and badly needed automation into the routine defense.

3.2.1. Transaction Narrative Analysis

One of the earliest and most effective uses of LLMs in finance is the analysis of transaction narratives, textual descriptions attached to payment transfers or financial logs. Traditional systems often fail to capture suspicious patterns hidden in benign-looking descriptions [19]. Fine-tuned BERT or RoBERTa-based models, however, will be able to analyze the semantics of the transaction notes, identify contradictions, and mark those phrases likely to be used in fraudulent or laundering transactions. These models offer a higher-level risk scoring granularity, in particular, for peer-to-peer, or high-volume transactions bulk payment, and on the corporate side.

3.2.2. Real-Time Support for SOC Teams

LLMs have also been deployed to assist Security Operations Center (SOC) analysts by automating routine tasks and enhancing decision support. Large autoregressive models such as GPT-2 and GPT-3 have been integrated into internal chatbots, enabling natural language querying of system logs, rapid summarization of incident reports, and contextual generation of response recommendations. In environments overwhelmed by alert fatigue, LLMs can synthesize incoming threat data into concise, actionable insights, thereby improving analyst efficiency and reducing mean time to respond (MTTR).

3.2.3. Phishing Detection Across Email, Voice, and Text

Phishing has become one of the most widespread attack vectors in the financial sector where the most advanced techniques are used with elements of social engineering, impersonation, and using multiple languages. LLMs have also gone a long way in identifying phishing across several forms of communication [20]. Even when adversaries hide malicious intent in obfuscation or when they adjust their communication behavior to look genuine, fine-tuned variants of BERT and T5 can detect manipulative intent in email and messages. In voice-service environments, large language models paired with automatic-speech-recognition engines listen to calls and translate spoken words into text on the fly. Working together, these tools spot voice-phishing scams by marking rehearsed lines, mismatched emotions, and tell-tale impersonation signals-a skill set far richer than what older auditing software can offer.

Table 2. Comparative Analysis of LLM Architectures Used in Financial Security Use-Cases

Model	Architecture	Strengths	Financial Security Use-Cases
BERT	Bidirectional Transformer	Strong in classification and contextual analysis	Phishing email detection, transaction log review
RoBERTa	Optimized BERT variant	Improved robustness and speed	Suspicious narrative parsing, anomaly detection
GPT-2	Transformer decoder	Fluent language generation	Alert explanation, chatbot responses
GPT-3	Large autoregressive model	Few-shot learning, strong contextual retention	SOC chatbot assistance, summarization of threat reports
T5	Text-to-text transfer	Versatile across generative and classification tasks	Policy translation, threat intelligence summarization

By enhancing the capacity of financial systems to comprehend and react to textual and linguistic threats, LLMs are proving essential in modern security infrastructures. These models provide dynamic and adaptive defenses as their competitors employ increasingly advanced narrative manipulation and it is critical that they no longer detect patterns but are semantically aware.

4. Federated Learning (FL) for Fraud Detection

As banks and payment firms lean ever more on artificial intelligence to spot fraud, worries about privacy and security have grown-intensifying whenever sensitive material such as transaction logs or behavioral biometrics enters the picture [21]. In response, Federated Learning (FL) has appeared as an attractive paradigm that allows carrying out model training in a distributed fashion across far-flung data sources without requiring centralization or raw data directly exchange. Such decentralized solution goes with the current regulatory and ethical pressure to minimize data storage and with the advantages of using data to detect fraudulent activity.

4.1. Motivation for FL

Traditional centralized machine learning implementations need all the training data to be brought to a centralized point and hence is very risky in the financial sector. Sensitive information such as an account activity, biometric patterns and payment histories can be compromised in the course of transmission and storage. In contrast, FL enables model training to occur locally on edge devices or institutional servers, sending only model updates (e.g., gradients) to a central aggregator. Because raw personal data never leaves its original location, exposure risk drops sharply and so does the likelihood of breaching GDPR or other financial privacy rules. The financial sector, where confidentiality is paramount, has thus become one of the early adopters of FL technologies. In scenarios such as fraud detection on mobile banking apps or biometric verification systems, FL provides a method to continuously update models with fresh, distributed data—without ever transmitting that data outside the user's device or institutional boundary.

4.2. FL Use Cases in Financial Security

4.2.1. Edge-Device Model Training for Fraud Detection

FL has been integrated into mobile banking environments to facilitate on-device learning. Each user's mobile app can locally train fraud detection models based on their unique behavioral patterns—such as login times, geolocation consistency, or touch dynamics without transmitting this data to a centralized server. Following a brief training period, the new weights are encrypted, transmitted back and added to the central server, reinforcing the shared model but leaving sensitive habits confidential. This makes the system learn quicker and respond earlier to new scams that exploit known patterns.

4.2.2. Cross-Bank Collaborative Models without Data Sharing

Another critical application involves inter-bank collaboration in model development without requiring direct data exchange. Traditionally, there are legal and competitive obstacles that block bank institutions to share customer records even when collaborative fraud detection is likely to be advantageous [22]. FL allows participating institutions to collaboratively train powerful models of fraud detection as they provide encrypted model updates based on their own confidential dataset. Such a configuration promotes sharing of threat intelligence across the industries and the robustness of these models in the detection of fraud patterns where the dimensions of the fraud might cross institutions or boundaries.

Figure 2. Federated vs. centralized model accuracy in detecting financial fraud, highlighting FL's privacy-preserving trade-offs.

Figure 2 illustrates the comparative results of federated and centralized learning models with a simulated dataset of fraud. Although centralized ones can have a little higher initial accuracy since their communication is based on access to complete datasets, federated models show competitive results with minimal privacy risks. The number demonstrates the feasibility of FL in practical financial conditions, particularly those in which privacy is important besides the precision of the models.

4.3. FL Frameworks Used

Several open-source frameworks have facilitated the practical adoption of FL in the financial sector. Google TensorFlow Federated (TFF) is a scalable interface with which it is possible to simulate and deploy federated learning workflows. It facilitates safe aggregation and different control measures, which implies that it can be applied in financial scenarios that require high privacy levels. PySyft, OpenMined-community maintained, emphasizes on secure multi-party computation and privacy budgets and can enable multiple banks or edge devices to share models without any raw data ever leaving their walls. Collectively, these platforms have fuelled prototypes and production systems of fraud detectors, identity verifiers and regulatory reporters. Federated Learning offers a privacy-preserving alternative to centralized data modeling that is well-suited for the financial industry. By enabling localized learning and secure collaboration, FL aligns technical innovation with regulatory requirements, making it an essential component in the evolving architecture of AI-driven financial security.

5. Graph Neural Networks (Gnns) For Transaction Behavior Modeling

Within any financial system, user actions and transaction flows tangle together in ways that simple lists rarely capture. Graph Neural Networks (GNNs) offer a natural tool kit in doing this since nodes become accounts, users, and devices while edges can be the transaction or connections between them. Instead, most typical machine-learning models assume that each observation is independent, but a GNN learns not about the form of the network, but rather picks up context about domains its neighbors inhabit; which makes the technique particularly useful at detecting hidden collusion or fraud that traverses multiple accounts.

5.1. Graph-Based Modeling of Financial Transactions

Financial ecosystems naturally exhibit graph-like characteristics. An individual user may be connected to multiple accounts, devices, or merchants, forming a web of interactions. Such interactions are generally dynamic and non-linear in nature and hence these kinds of interactions are difficult to model and interpret in terms of flat or tabulation of data. GNNs do not have this restriction as they can use nodes in a graph and pass the messages to them and each node can combine what it gets in vicinity and update itself accordingly [23]. Graph-based learning stands out in fraud prevention because suspicious actions seldom occur in isolation. One transaction might not look untoward, but there can be matryoshka nesting of transactions with others that wash out its legitimacy. GNNs thus enables analysts to measure the local neighborhood as well as the entire network structure, to provide a more well-balanced view of each account and transfer.

5.2. Applications

5.2.1. Fraud Ring Detection

Some promise of GNNs has been demonstrated with respect to detecting fraud rings (i.e. group of interconnected accounts committed to collusive activity). Such rings would tend to escape capture by more conventional theories as they are distributed and are on the subtle side of interaction. By analyzing multi-hop relationships and cyclic graph structures, GNNs can detect hidden communities indicative of orchestrated fraud, improving both precision and recall in complex attack scenarios.

5.2.2. Community Detection and Anomalous Behavior

GNNs also serve broader community detection goals, sorting nodes with similar behavior into coherent groups. When a member suddenly deviates-louder spending bursts, login from new devices, or mismatched identity clues-the system flags it as a possible breach. Models such as GraphSAGE or domain-specific heterogeneous nets for example are great at learning these patterns into compact embeddings to keep unsupervised track and intervene in time to prevent larger losses.

5.2.3. Transaction Path Scoring

In transactional graphs, the sequence and structure of connections matter. GNNs can also rate the transaction paths by examining how money flows to the network. As an example, a payment that takes place between a chain of highly risky or lightly connected nodes could be an attempt at laundering. GNNs allow construction of higher degree of accuracies and context-sensitive fraud detection models due to the consideration of temporal and structural dependencies.

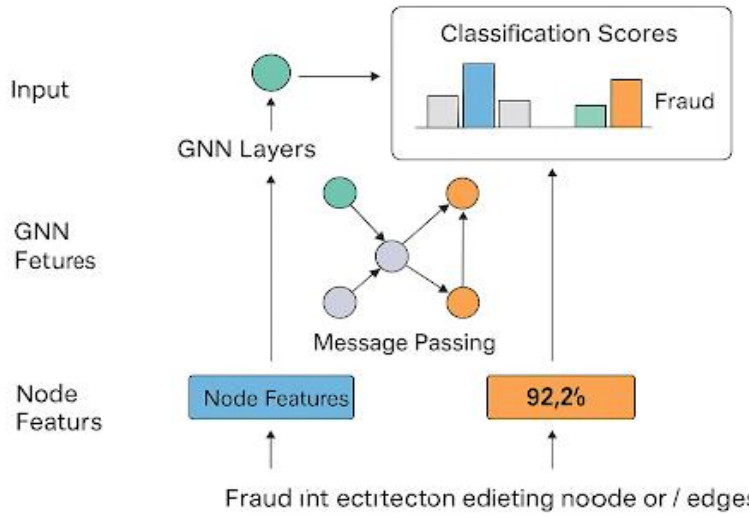

Figure 3. GNN-based model for detecting fraud in transaction networks.

Figure 3 illustrates a standard GNN structure when used on a transaction network. Considering node and edge as input features, the model passes the neighborhood information through the message passing mechanism, and provides the output in classification scores of a possible fraud node or an edge.

5.3. Notable Models

Several GNN architectures have been adopted for financial applications, each offering specific advantages depending on graph complexity and data characteristics:

5.3.1. Graph Convolutional Networks (GCN)

Suitable for static, homogenous graphs with well-defined neighborhoods. Commonly used in structured financial datasets like blockchain graphs.

- **GraphSAGE:** It can easily add new nodes by learning examples on the fly, thereby well suited in dynamic services such as internet banking or peer-to-peer payment applications.
- **Graph Attention Networks (GAT):** Employ attention mechanisms to assign different weights to neighboring nodes, enhancing performance in graphs with noisy or unbalanced structures.
- **Heterogeneous GNNs:** It handle many node and edge types at the same time and represents the sprawling network of users, devices, locations, merchants and so on that a modern financial ecosystem is flooded with.

Table 3 presents a benchmark comparison of popular GNN models evaluated on financial datasets, demonstrating their effectiveness in tasks like fraud classification, risk scoring, and user profiling.

Table 3. Different GNN Models

Model	Task	Key Strengths
GCN	Bitcoin fraud detection	Efficient in dense transaction graphs
GraphSAGE	Identity fraud detection	Generalizes to unseen users/accounts
GAT	Transaction risk scoring	Learns importance of neighbors
Hetero-GNN	Multi-entity fraud detection	Captures complex entity relationships

By treating financial data as an evolving graph rather than isolated records, GNNs offer enhanced contextual intelligence, leading to more accurate and explainable fraud detection outcomes. They are versatile to various financial systems and can be used as a technology base to support future much more powerful AI-assisted transaction monitoring.

6. Behavioral Biometrics for Continuous Authentication

Behavioral biometrics has emerged as an essential backup to classic authentication tools in finance. Instead of relying on one-off secrets like passwords or PINs,it closely monitors the minute details of the habits of a user with their device and can get their identity confirmed all through a session silently. This monitoring is passive and continuous, and is therefore a secure method of tightening security and allowing smooth movement of customers, which makes the method perfect in detecting fraud and preventing identity in habitations and payments.

6.1. Overview of Behavioral Biometric Modalities

Behavioral biometric systems monitor patterns that are difficult to replicate or forge, even by sophisticated attackers. Common modalities include keystroke dynamics (typing rhythm, key press duration, and latency), mouse movements, touchscreen gestures, devices grip patterns, and gait patterns with embedded sensors. Upon being recorded at consecutive instances, these behavioral indicators present exceptional biometric signatures that like one another but are still discernible and change gradually per person. Unlike fingerprints or facial scans, which measure fixed body features, behavioral channels are always aware of their surroundings and can shift as circumstances change[24]. They show how a person adapts to different apps, devices, and lighting, so they resist attack methods like video replay, molded masks, or stolen photos. Since the analysis can occur on smartphones, laptops or even smart payment terminals, the analysis can basically be device-neutral, and thus, businesses won t have to reinvent the wheel to protect various platforms.

6.2. Integration with AI Models

Artificial Intelligence, especially machine learning devices such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) are very important in getting meaningful representations out of behavioral data. RNNs are well-suited for modeling sequential inputs such as keystroke timings or swipe gestures, while CNNs are employed to capture spatial patterns in touch heatmaps or pressure points[25]. It is common to pair these architectures with anomaly detection where the system will be able to update the normal pattern of each user in real time and highlight any perceptible deviation to be investigated further. The easiest one to list an example of a banking app: in case the typing pattern of a person changes in stressful situations or as the result of an apparent hacking attempt, the service may begin additional verification procedures or shut down the session automatically. This ongoing artificial intelligence-driven supervision reduces strain on one-time-shot logon passwords and provides organizations with an early warning line of defense.Table 4 below presents a comparative overview of behavioral biometric modalities, their typical applications, and effectiveness in financial security contexts.

Table 4. Behavioral Biometric Modalities and Their Effectiveness in Financial Applications

Modality	Common Use Case	Detection Focus	AI Technique Used	Effectiveness (Reported)
Keystroke Dynamics	Web-based logins	Account takeover attempts	RNNs	High
Touch Dynamics	Mobile banking apps	Synthetic identity usage	CNNs	High
Mouse Movements	Online transaction validation	Bot detection	Hybrid models	Moderate
Device Grip	Smartphone payments	Session hijacking	RNNs	Moderate–High
Gait Recognition	Wearables and mobile platforms	Continuous authentication	RNNs/CNNs	Emerging

6.3. Deployment Examples in Financial Applications

Leading banks and security vendors have already rolled out behavioral biometric tools with clear success. Noteworthy solutions include:

- **BioCatch:** A pioneer in the field, this provider reviews more than 2,000 interaction traits to flag online fraud in real time.
- **Zighra:** Designed for mobile devices, its on-phone machine learning delivers fast, low-power authentication without noticeable delay.
- **BehavioSec:** Focuses on enterprise-grade behavioral biometrics, integrating seamlessly with existing identity and access management systems.

These systems have demonstrated measurable improvements in fraud detection rates, particularly in identifying account takeover attempts and synthetic identity misuse, without introducing user friction.

7. Privacy, Ethics, and Regulatory Considerations

As banks and payment platforms increasingly lean on artificial intelligence to safeguard transactions, new questions arise about privacy, compliance with the law, fairness in algorithms, and the responsible handling of personal information. Technologies such as large language models, federated learning, graph neural networks, and behavioral biometrics can noticeably boost fraud detection, yet their use must still weigh strong security against clear reporting, sound data safeguards, and the publics confidence.

7.1. Data Privacy Regulations and Their Implications

AI-driven fraud detection systems often process sensitive financial and biometric data that fall under the jurisdiction of major data privacy regulations.General Data Protection Regulation (GDPR) in the European Union proposes the principles of data minimization, limit of purpose, and user authorization that directly affect the collection process, storage, and usage of the training data. On the same note, PSD2 (payment services directive 2) introduces strong customer authentication (SCA) and establishes security innovation in a controlled environment[26]. In the United States, California Consumer Privacy Act (CCPA), which provides consumers even a direct right of accessing their data, deleting it, and the right to opt out of the collection entirely. This legal transition pressurizes the existing opaque AI systems that monitor behavior in the long term and by extension, pressure on financial institutions that are using them. The requirement to respect privacy has made banks and fintech companies turn to privacy-first algorithms like differential privacy, homomorphic encryption, and federated learning, to ensure their AI does not raise consumer suspicion and expose them to lawsuits.

7.2. Explainability and Algorithmic Bias in AI Models

One of the most pressing ethical challenges is the lack of explainability in high-capacity AI models like LLMs and GNNs. While these models are powerful, their decision-making processes often remain opaque, raising concerns when used in high-stakes environments such as fraud risk assessment or identity verification. Explainable AI (XAI) capability is therefore no longer optional; regulators are setting it as a requirement, and boards insist on it as a deployment dependent precondition. Machine-learning tools like SHAP (SHapley Additive explanations), LIME (Local Interpretable Model-agnostic Explanations), and attention heat maps convert abstract features to down-to-earth narratives delivered to help auditors and compliance teams produce more accurate results. Bias is another ethical risk. AI systems trained on skewed or non-representative datasets may exhibit discriminatory behavior, such as falsely flagging certain user demographics as high-risk. Addressing this requires bias audits, diverse training data, and fairness constraints embedded directly into model objectives. Figure 4 illustrates the AI Tradeoff Triangle on financial security- with the focus on the fact that security, privacy, and explainability are linked as opposite goals.

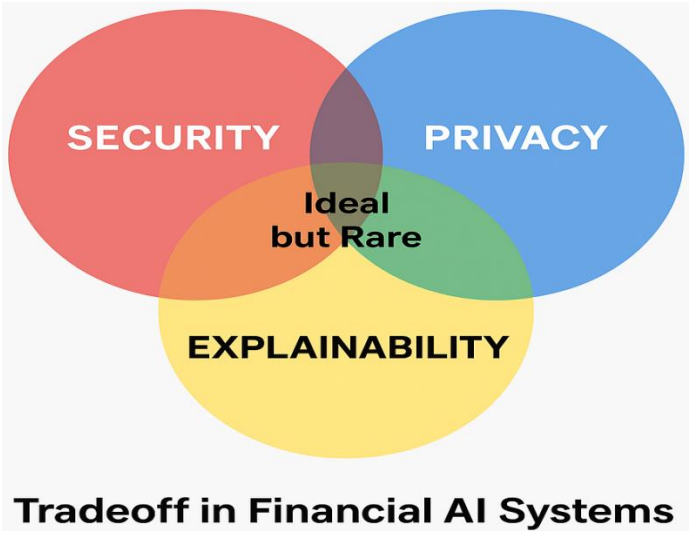

Figure 4. Tradeoff Triangle in Financial AI Systems: Balancing Security, Privacy, and Explainability.

7.3. Risks of Federated Attacks and Model Inversion

Even systems that seem privacy-friendly, such as federated learning, can fall prey to clever adversaries. With model inversion, a hacker is able to reverse-engineer important inputs- including log in particulars or latest transactions-via analysing the shared weights. Gradient leakage, membership inference, and other related attacks filter the updates posted by the local nodes to extract the privacy of habits or identities. Poisoning assaults add another layer of danger: a compromised smartphone or laptop feeds false examples into the network and slowly spoils the global model. Defenders can also restrict the destruction by employing secure aggregation protocols, integrating differential privacy in each move, and executing sensitive code in the so-called trusted execution environments (TEEs) when passing on updates. Protecting decentralized AI platforms from hostile intrusions thus remains vital if the privacy promises of those systems are to hold. Regulators around the world are growing more aware of the issue and may soon demand tougher proof of adversarial strength from any AI unit serving finance.

8. Comparative Evaluation

Arrival of large language models (LLM), federated learning (FL), graph neural networks (GNNs), and behavioral biometrics has taken fraud detection to a whole new level, leavingdated, rule-based engines and predetermined feature sets in the back seat. These AI-first gambits are contrasted in the section below with the conventional approach in terms of accuracy, false positive rate (FPR), speed of detection, scalability, and privacy in their regard to them. Established systems usually rely on manually written rules, fixed cut-offs, or lightweight classifiers such as decision trees or logistic regression. These models are both fast to operate and simple to describe, but they cannot stay abreast of changing strategies of fraud hence produce loads of false flags and require infinite human modifications. Theirs is a short perspective on every transaction and cannot draw larger charts of their patterns of purchases, and so are helpless against rings, synthetic identities, or phishing schemes that operate in the gray.

The LLMs also have the capacity to read and parse unstructured communications in the financial sphere and online phishing or social engineering attempts, and the narratives of transactions. Federated Learning allows training of a model on distributed banks or mobile devices without needing to centralize sensitive data, making it able to comply with privacy regulations. The strength of GNNs is that they identify structural abnormalities within the networks of transactions, discovering undisclosed connections between malicious individuals. Behavioral biometrics When combined with neural networks, behavioural biometrics can offer continuous, frictionless user authentication on the basis of user interaction patterns. Table 5 below presents a consolidated view of how these AI techniques perform across various evaluation metrics relevant to financial security applications.

This evaluation shows that while traditional systems may still offer value in specific low-risk scenarios due to their simplicity, they fall short in the face of modern fraud vectors that demand adaptability, contextual reasoning, and cross-platform intelligence. Among the AI approaches, federated learning and GNNs particularly stand out for their ability to detect subtle fraud patterns while maintaining user privacy. However, latency and deployment complexity remain challenges for large-scale implementations. Overall, a hybrid deployment strategy, combining rule-based logic and AI models could well provide the best of both worlds in the sense of allowing customers to interpret the model, but take advantage of the adaptive capabilities of a deep learning system. Fractional analysis: These trade-offs will need to consider the level of risk tolerance of the organization, regulatory requirements, and the level of technical maturity financial institutions must address when considering modernization of its fraud detection infrastructure.

9. Challenges and Limitations

While AI-driven techniques have significantly advanced the detection and prevention of fraudulent financial activities, their deployment in real-world environments remains fraught with challenges. operational and systemic levels, and are associated with the long-term effectiveness and sustainability of models. It is important that practitioners and researchers understand such barriers to maximize the AI-based solutions in the financial contexts.

9.1. Data Scarcity and Imbalance

Effective AI models rely on large volumes of high-quality, labeled data. In financial fraud detection, this requirement is often hindered by data scarcity, particularly for rare or novel fraud cases. Fraudulent transactions typically constitute a small fraction of all financial activities, leading to high class imbalance. This skews learning algorithms toward the majority class (legitimate transactions), reducing sensitivity to minority class anomalies. Even advanced architectures like GNNs or LLMs may underperform without adequate representation of evolving fraud patterns. Techniques such as oversampling, SMOTE, and cost-sensitive learning partially address this, but the fundamental challenge of limited ground truth persists.

9.2. Real-Time Processing Constraints

In real-time fraud detection scenarios—such as those encountered by financial institutions or retail systems—decision-making must occur with minimal latency. However, many state-of-the-art models, including large-scale transformers and graph neural networks, are computationally intensive and memory-demanding, thereby introducing delays that can erode user trust. The need for low-latency inference imposes significant constraints on the deployment of such complex architectures in time-sensitive environments. This challenge is particularly pronounced in edge devices such as mobile phones or point-of-sale gateways, which typically possess limited processing capabilities. As a result, there is a growing preference for lightweight or quantized models that strike a balance between inference speed and predictive accuracy. Achieving this balance remains a central concern for development teams working with large language models or multimodal biometric systems in operational settings.

9.3. Model Drift and Fraud Evolution

AI-based fraud detection systems are inherently vulnerable to concept drift, wherein the statistical properties of input data evolve over time. Since these models are often trained on historical datasets, their performance may degrade as adversaries adapt their techniques in response to deployed detection mechanisms. Attackers frequently employ strategies such as generating adversarial examples or exploiting rare patterns not adequately represented in the training data. Consequently, models that were once highly effective may become obsolete unless they undergo continuous retraining and evaluation. To maintain detection efficacy, organizations must implement adaptive learning pipelines capable of integrating recent data and deploying model updates rapidly and efficiently.

Table 5. Comparative Performance Metrics of Different AI-Based Techniques

Technique	Accuracy	False Positive Rate	Latency	Scalability	Privacy Compliance
Rule-Based Systems	Moderate	High	Low	Limited	Basic
LLMs (e.g., BERT, GPT)	High	Low–Moderate	Moderate	High (with tuning)	Moderate
Federated Learning	High	Low	Moderate–High	High	High
Graph Neural Networks	High	Low	Moderate	Moderate	Moderate
Behavioral Biometrics	Moderate	Low	Low	High	High

9.4. Federated Learning Communication Overhead

While Federated Learning (FL) addresses privacy and data locality concerns, it introduces communication overhead between client devices and the central aggregator. Frequent model updates, particularly in large networks of mobile devices or cross-bank collaborations, strain bandwidth and increase synchronization delays. Additionally, heterogeneous data distributions across clients (non-IID data) may reduce model convergence speed and accuracy. Solutions like model compression, asynchronous updates, and personalized federated learning have been proposed, yet communication inefficiency remains a barrier to the widespread adoption of FL in latency-sensitive financial systems.

10. Conclusion

The integration of artificial intelligence into financial security systems has marked a significant paradigm shift in how institutions detect, analyze, and respond to fraudulent activity. By leveraging cutting-edge approaches such as Large Language Models (LLMs), Federated Learning (FL), Graph Neural Networks (GNNs), and behavioral biometrics, the financial industry is gradually moving toward more intelligent, context-aware, and privacy-preserving security frameworks. These technologies enable not only higher detection accuracy but also the ability to adapt to evolving threat patterns and deliver more seamless user experiences. Large language models (LLMs) now power advanced language understanding, letting them read transaction notes on the fly and flag phishing in email, chat, and SMS streams. Federated learning (FL) walks the tightrope between performance and privacy by letting banks train a shared model without sharing raw customer data. Graph neural networks (GNNs) then map the connections between accounts and devices, revealing tangled fraud rings that standard rule engines

usually miss. Add in behavioral biometrics-passive keystroke, touch, or mouse movement checks-and firms gain seamless security that hardly interrupts a legitimate user.

However, the adoption of these AI technologies is not without limitations. Data imbalance, real-time processing constraints, and the dynamic nature of fraud introduce persistent challenges. Federated learning systems face significant communication overhead and require optimization to perform effectively at scale. Moreover, the ethical and regulatory landscape shaped by frameworks such as GDPR, PSD2, and CCPA, demands that AI systems remain interpretable, auditable, and fair, especially when applied to sensitive domains like finance. The comparative analysis presented in this review underscores that no single artificial intelligence technique consistently outperforms others across all fraud detection scenarios. Consequently, the adoption of stacked or hybrid models where multiple algorithms operate in a modular and complementary manner emerges as a robust and versatile approach to countering the diverse and rapidly evolving threats within the financial domain. Future research should prioritize enhancing model interpretability, optimizing architectures for real-time deployment, and developing adaptive systems capable of learning from limited, imbalanced, or adversarial datasets. Ultimately, the intersection of AI and financial security represents both an opportunity and a responsibility. When carefully designed and ethically deployed, these technologies hold the potential to redefine trust, safety, and transparency in the global financial ecosystem.

References

[1] P. Gomber, J. Kauffman, C. Parker and B. Weber, "On the fintech revolution: Interpreting the forces of innovation, disruption, and transformation in financial services," J. Manage. Inf. Syst., vol. 35, no. 1, pp. 220–265, 2018.

[2] S. Dixit and J. Jangid, "Exploring smart contracts and artificial intelligence in FinTech," J. Inf. Syst. Eng. Manage., vol. 10, no. 14s, pp. 282–295, 2025. [Online]. Available: https://doi.org/10.52783/jisem.v10i14s.2208

[3] M. Bojilov, Methods for assisting in detection of synthetic identity fraud in credit applications in financial institutions, Ph.D. dissertation, CQUniversity, 2023.

[4] S. Dixit, "The impact of quantum supremacy on cryptography: Implications for secure financial transactions," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 6, no. 4, pp. 611–637, 2020. [Online]. Available: https://doi.org/10.32628/CSEIT2064141

[5] N. S. Al-Musib, et al., "Business email compromise (BEC) attacks," Mater. Today: Proc., vol. 81, pp. 497–503, 2023.

[6] V. Mahalakshmi, et al., "The role of implementing artificial intelligence and machine learning technologies in the financial services industry for creating competitive intelligence," Mater. Today: Proc., vol. 56, pp. 2252–2255, 2022.

[7] F. Ahmed, "Quantum-resistant cryptography for national security: A policy and implementation roadmap," Int. J. Multidiscip. Sci. Manage., vol. 1, no. 4, pp. 54–65, 2024.

[8] A. Winograd, "Loose-lipped large language models spill your secrets: The privacy implications of large language models," Harv. J. Law Technol., vol. 36, pp. 615, 2022.

[9] L. Bellomarini, E. Laurenza, and E. Sallinger, "Rule-based anti-money laundering in financial intelligence units: Experience and vision," in Proc. RuleML+ RR (Suppl.), vol. 2644, pp. 133–144, 2020.

[10] P. M. Mammen, "Federated learning: Opportunities and challenges," arXiv preprint, arXiv:2101.05428, 2021.

[11] P. Kałużny, "Behavioral biometrics in mobile banking and payment applications," in Proc. Int. Conf. Bus. Inf. Syst., Cham: Springer, 2018.

[12] I. de Zarzà, et al., "Optimized financial planning: Integrating individual and cooperative budgeting models with LLM recommendations," AI, vol. 5, no. 1, pp. 91–114, 2023.

[13] J. Nicholls, A. Kuppa, and N.-A. Le-Khac, "Financial cybercrime: A comprehensive survey of deep learning approaches to tackle the evolving financial crime landscape," IEEE Access, vol. 9, pp. 163965–163986, 2021.

[14] K. C. Aguoru, An empirical investigation of the causes and consequences of card-not-present fraud, its impact and solution, Ph.D. dissertation, Univ. East London, 2015.

[15] N. Zachosova and N. Babina, "Identification of threats to financial institutions' economic security as an element of the state financial security regulation," Baltic J. Econ. Stud., vol. 4, no. 3, pp. 80–87, 2018.

[16] Y. Li, et al., "Large language models in finance: A survey," in Proc. 4th ACM Int. Conf. AI Finance, 2023.

[17] S. Wu, et al., "Bloomberggpt: A large language model for finance," arXiv preprint, arXiv:2303.17564, 2023.

[18] O. Campesato, Transformer, BERT, and GPT: Including ChatGPT and Prompt Engineering, 2023.

[19] J. Vos and B. van Rijn, "The evidence-based conceptual model of transactional analysis: A focused review of the research literature," Transact. Anal. J., vol. 51, no. 2, pp. 160–201, 2021.

[20] F. Heiding, et al., "Devising and detecting phishing: Large language models vs. smaller human models," arXiv preprint, arXiv:2308.12287, 2023.

[21] W. Yang, et al., "FFD: A federated learning based method for credit card fraud detection," in Proc. BigData 2019: 8th Int. Congr. as part of SCF 2019, San Diego, CA, USA, Jun. 25–30, 2019.

[22] T. Suzumura, et al., "Towards federated graph learning for collaborative financial crimes detection," arXiv preprint, arXiv:1909.12946, 2019.

[23] J. Wang, et al., "A review on graph neural network methods in financial applications," arXiv preprint, arXiv:2111.15367, 2021.

[24] Eberz, et al., "Evaluating behavioral biometrics for continuous authentication: Challenges and metrics," in Proc. ACM Asia Conf. Comput. Commun. Security, 2017.
[25] A. Dalsaniya, "AI for behavioral biometrics in cybersecurity: Enhancing authentication and fraud detection," Int. Res. J., vol. 10, pp. a108–a122, 2023.
[26] J. Jangid and S. Dixit, The AI Renaissance: Innovations, Ethics, and the Future of Intelligent Systems, vol. 1, Technoscience Academy, 2023.