International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(REVIEW ARTICLE)

Artificial Intelligence and machine learning in fraud detection for digital payments

Alexandre Davitaia *

Independent Researcher, Germany.

International Journal of Science and Research Archive, 2025, 15(03), 714-719

Publication history: Received on 28 April 2025; revised on 08 June 2025; accepted on 10 June 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.15.3.1784

Abstract

The financial sector has adopted artificial intelligence (AI) and machine learning (ML) for more advanced and real-time fraud detection as a result of the increased threat of fraud brought on by the global surge in digital payments. By using sophisticated algorithms like supervised learning, anomaly detection, and deep neural networks, these technologies allow systems to identify irregularities, adjust to new fraud patterns, and lower false positives. AI-driven systems are widely used by fintechs and neobanks in the US, Germany, and the EU. However, they also face issues with data quality, model transparency, and regulatory compliance, which calls for a careful balancing act between ethical oversight and technical solutions.

Keywords: AI Digital Payments; AI Fraud Detection; AI transaction detection; AI in fintech

1. Introduction

Trillions of transactions are processed each year thanks to the explosive growth of digital payments. In 2023, for instance, the global payments sector processed 3.4 trillion transactions totaling $1.8 trillion. Fraud attempts increase in tandem with transaction volumes; according to one source, the number of suspected financial services fraud cases increased by 39% between 2019 and 2022. The scope and complexity of contemporary fraud are too great for conventional rule-based systems (such as velocity checks or fixed fraud rules). In response, banks, fintechs, and regulators increasingly turn to AI/ML for real-time, adaptive fraud detection. Large data streams can be analyzed by ML models, which can then identify subtle patterns and anomalies that point to fraud. By continuously learning from new cases, these models can increase accuracy and decrease false alarms. Industry observers call this shift "game-changing": AI tools are viewed as essential to *reduce false positives and improve accuracy* in fraud prevention. However, deploying AI in payments also raises technical and policy challenges (e.g. bias, privacy, interpretability), which we discuss below.

2. Machine Learning Techniques and Models

Fraud detection uses a variety of machine learning techniques. In actuality, supervised learning predominates; researchers discovered that supervised techniques were employed in roughly 56.7% of fraud detection studies, while unsupervised techniques were employed in about 18% of them. Algorithms are trained on transaction data classified as "fraud" or "legitimate" in supervised models. Neural networks, logistic regression, decision trees, random forests, support vector machines, k-nearest neighbors, and naive Bayes are typical examples. These classifiers use historical data to predict the likelihood that a new transaction will be fraudulent.

By contrast, unsupervised and semi-supervised learning is used when labels are scarce. Unsupervised models analyze transaction data without prior fraud labels to detect *anomalies*. For instance, clustering methods (e.g. DBSCAN) group typical transaction behavior so that outliers stand out. Autoencoders (deep neural nets) can be trained to reconstruct normal transactions; large reconstruction errors signal anomalies. Isolation Forests repeatedly partition the feature

* Corresponding author: Alexandre Davitaia

space to isolate data points – anomalies tend to be isolated quickly and receive high *anomaly scores*. In one explanation: *"Isolation forest attempts to isolate anomalies as the first step. … Each point receives an anomaly score; values above a threshold are more likely to be anomalous"*. These unsupervised models can flag novel fraud patterns (e.g. a new scam tactic not seen before). Hybrid approaches also exist: semi-supervised models or ensembles that combine supervised and unsupervised outputs can leverage both labeled fraud examples and clustering of unlabeled data.

Deep learning (deep neural networks) is also emerging, especially for complex data (e.g. analyzing sequences of user behavior or raw text). In practice, deep learning is a small percentage of existing deployments (~2–3% of studies) but growing. Convolutional or recurrent networks could, for example, learn intricate patterns in sequential transaction streams or device fingerprints. In all cases, models output a risk score for each transaction; if the score exceeds a threshold, an alert is raised.

Key ML techniques in fraud detection include:

- Classification algorithms (supervised): LR, SVM, decision trees, random forests, neural networks (trained on labeled fraud cases).
- Anomaly detection (unsupervised): Clustering (DBSCAN), autoencoders, isolation forests (score outliers).
- Hybrid methods: Semi-supervised models combining limited labels with unlabeled data, or ensembles of supervised/unsupervised models.
- Real-time scoring: Streaming ML systems that process each transaction (or batch) in real time using trained models and risk rules (see next section).

These techniques are often supplemented with feature engineering: e.g. deriving historical spending profiles, device identifiers, location context, network graphs of accounts, etc. Modern fraud detection also uses graph-based methods (graph neural networks) to detect networks of colluding accounts or money mules.

3. Model Deployment and Real-Time Monitoring

Deploying fraud models typically involves real-time or near-real-time pipelines. Fraud detection is time-critical: once a suspicious payment is initiated, banks must block or flag it *immediately* to prevent loss. Thus, many systems attach ML models directly to transaction processing. For example, cloud-based services or on-premises engines evaluate each payment request on the fly, combining ML scores with business rules. Real-time model deployment requires low latency and high reliability; some firms use stream-processing platforms (e.g. Apache Kafka/Spark) or specialized in-memory databases.

AI-driven systems continuously learn and adapt. Unlike static rule engines, ML models can be retrained on fresh data so they evolve as fraud tactics change. Seth K notes that adaptive machine learning models … learn from historical fraud cases and genuine transactions. Over time, their accuracy continuously improves, reducing the chances of flagging legitimate transactions as fraudulent. In practice, banks maintain pipelines to periodically retrain models on new labeled cases (including confirmed frauds) and even use synthetic data or simulations to cover rare scenarios. Continuous training reduces "concept drift" when fraud patterns shift.

A key benefit of AI-based deployment is real-time decision-making. As one industry report explains, modern fraud systems leverage deep learning and neural nets so that "AI models assess risk instantly, reducing the reliance on manual reviews. This improves efficiency and enhances customer experience by preventing unnecessary transaction declines". For instance, the system can dynamically allow a legitimate high-value overseas purchase (based on customer history and context) while blocking subtle yet fraudulent patterns (e.g. a new device or unusual merchant). Rapid model scoring often runs in milliseconds on each transaction.

Key deployment features include:

- Streaming scoring: Models integrate into payment flows to score each transaction in real time.
- Continuous learning: Automated retraining on new fraud cases (possibly with human-in-the-loop) to update model parameters.
- Behavioral analytics: Systems track user/device behavior (habits, geolocation, device IDs) so alerts have contextual rules.
- Explainability (XAI): Modern solutions incorporate explainable AI tools so investigators can understand *why* a transaction was flagged. For example, clear rule traces or feature contributions (e.g. "unusual location" triggered the alert) help satisfy audit and regulatory requirements. Seth K notes that XAI provides

clear insights into why a transaction was flagged, ensuring compliance with regulatory requirements while reducing false positives. This transparency is crucial for regulators and customer appeals.(Seth, 2025)

In summary, deployment combines sophisticated ML engines with fraud analysts and automated workflows. Many organizations now employ AI-based decision systems: if a payment's risk score exceeds a threshold, it is blocked or held for review; otherwise it is auto-approved. This markedly reduces the workload on manual fraud teams and increases detection speed.

4. Industry Applications and Examples

AI-powered fraud detection is widely adopted across the payments ecosystem. Fintech firms and neobanks are leaders in this area, since their business is 100% digital. For example, *digital banks like N26 (Germany), Monzo and Starling (UK), and fintechs like Revolut and Wise* heavily leverage ML for fraud screening. These online platforms use AI in several ways: fraud scoring at account sign-up (KYC), continuous transaction monitoring, and real-time authentication (e.g. biometric login). Fintechs collect rich user data, enabling ML models to build detailed risk profiles. As one analysis notes, *"Many neobanks rely heavily on automation, artificial intelligence, and machine learning for fraud detection. While these technologies are powerful, they are not infallible"*. These firms often use ML-driven device analysis (e.g. linking devices to accounts) and network analysis to spot account takeovers, transaction anomalies, and money mule schemes.

Traditional banks and payment networks have also integrated AI. For instance, *Visa's Decision Manager* is an automated fraud engine used by banks and merchants: it "calculates a risk value for each transaction using artificial intelligence and specified rules," drawing on global Visa network data. Deutsche Bank partnered with Visa to offer this to its merchants, explaining that the solution uses *"advanced risk models and global data intelligence from billions of data points … allowing good transactions to be accelerated and suspected fraudulent transactions to be blocked"*. (Deutsche Bank, 2022) By 2021, Visa reported that Decision Manager prevented over $22 billion in fraud worldwide. Similarly, Mastercard and network partners continuously analyze authorization flows with AI to flag stolen cards or synthetic identity fraud.

In the U.S. financial system, major players like PayPal, Stripe, and banks deploy ML at scale. While corporate details are proprietary, it is known that PayPal's risk engine has used AI for years to scan purchase velocity and user patterns. Stripe's Radar product, for example, uses a machine learning model trained on data from all Stripe customers to detect fraud (though we did not directly cite it here). On the regulatory side, the U.S. Treasury has explicitly embraced ML: its Bureau of the Fiscal Service reported using machine learning to catch Treasury check fraud, resulting in $1 billion in fraud and improper payment recoveries. In FY 2024, Treasury announced that ML and AI enhancements "prevented and recovered over $4 billion" in fraud. These figures underscore how AI tools are being applied at government scale to secure payments and reimbursements.

In Europe, aside from N26 and Revolut, traditional banks are also upgrading. For example, Deutsche Bank now offers AI-driven fraud screening for e-commerce merchants (as noted above). Under PSD2 regulations, banks must also implement *transaction monitoring* for strong authentication; in practice, many banks use ML models to analyze payment attributes and device signals to fulfill this requirement. Emerging neobanks in Germany and France similarly use biometric and behavioral analytics (e.g. continuous selfie/face checks, location monitoring) to spot money mules and account takeovers.

Finally, payment processors and card networks provide AI services to merchants. Visa's Cybersource and Mastercard's Decision Intelligence, for example, apply ML across all their clients. They combine this with consortium data – fraud patterns seen across millions of cards – to detect fraud earlier. This collaborative data advantage means the AI models at networks benefit from a global view of fraud trends.

Overall, AI/ML in fraud detection spans fintech startups to legacy banks. Any organization processing transactions can integrate ML risk engines. The trend is clear: as fraud grows, firms across Europe and the U.S. are deploying these advanced analytics to stay ahead.

4.1. Policy and Regulatory Frameworks

AI fraud detection sits at the intersection of finance regulation and data law. Key frameworks shape how these systems are built and used:

- **Payment services regulations:** In Europe, the Revised Payment Services Directive (**PSD2**, 2018) mandates *strong customer authentication* and encourages continuous monitoring of transactions. For example, PSD2's technical standards explicitly call for "transaction and device monitoring to identify unusual payment patterns". This regulatory push has accelerated banks' adoption of analytics (often ML-based) to meet real-time monitoring requirements.

- **Data protection (GDPR):** Under EU **GDPR**, payment data is sensitive personal information. AI systems must comply with privacy rules (data minimization, purpose limitation) and often cannot freely share data across borders. As one analysis notes, *"Regulations like GDPR restrict data sharing across institutions, complicating collaborative fraud detection"*. In other words, while cross-institution intelligence would improve fraud spotting (e.g. sharing blacklisted identities), privacy laws can hinder it. Companies must therefore employ anonymization, federated learning (see below), or obtain clear consents. GDPR also gives users rights (e.g. the right to explanation) that can impact how automated decisions are justified.

- **Anti-Money Laundering (AML) directives:** Both the EU (AMLD4/5/6) and the U.S. Bank Secrecy Act require financial institutions to monitor for suspicious activity and report it. AI/ML is now an accepted tool for AML compliance. As the German regulator BaFin notes, big data and ML are *"typically used for anti-money laundering and fraud detection"* in banking. BaFin has even published guidance on algorithms in finance, emphasizing principles like clear management oversight, robust data governance, and rigorous validation of models. In practice, banks using ML must maintain auditable processes (often via explainable AI) to satisfy regulators.

- **AI ethics and governance**: New regulations unique to AI are also relevant. Transparency, accuracy, fairness, and human oversight are among the stringent requirements that the upcoming Artificial Intelligence Act (EU AI Act) will impose on high-risk AI systems in the EU, which is likely to include credit and fraud scoring. For high-risk models, for instance, it requires accuracy/bias assessments and risk management procedures. AI risk is also a concern for U.S. regulators. The U.S. Treasury (2024) stressed the need for high-quality data and fairness monitoring and suggested precise definitions and guidelines for the application of AI in finance. Some U.S. agencies (like NYDFS) have started to release guidelines at the state level regarding AI explainability in finance and model risk management.

- **Standards and industry practices:** Fraud systems are indirectly impacted by industry standards such as PCI DSS (for card data) in addition to the law. Additionally, efforts like the Financial Stability Board's consultations are encouraging international cooperation on AI. All things considered, businesses have to manage a complicated web of regulations, ranging from those pertaining to AI and anti-fraud to those pertaining to customer privacy. This frequently entails creating "compliance by design," such as encrypting data, recording model choices, and keeping up with emerging regulations.

4.2. Trends and Challenges

While AI brings benefits, it also introduces challenges.

False positives vs. accuracy: A classic problem in fraud detection is the trade-off between catching fraud and inconveniencing customers. Rule-based systems often flag many legitimate transactions (high false positives). AI helps reduce this: ML models, by learning nuanced patterns, can cut false alerts. For example, Seth k reports that AI screening *"tracks spending habits, transaction frequency, location patterns, and device usage"* to distinguish genuine behavior from fraud, and that deploying ML significantly minimized false positives in one case study. (Seth, 2025) . However, no model is perfect: overly aggressive AI thresholds can still wrongly block valid payments. Tuning models (thresholds, retraining) is an ongoing task. As one analyst put it, fraud detection now requires "juggling lowering false alerts while stopping actual fraud".

Bias and data quality: AI systems are only as good as their data. In fraud detection, training data is inherently imbalanced (frauds are rare) and may reflect historical biases. For example, certain countries, demographics, or behaviors might be over-represented in past fraud labels, leading models to unfairly target them. If unchecked, an ML system could discriminate by race, age, or location. Regulators (and researchers) are keenly aware of this. The EU AI Act and regulators like BaFin demand that firms address fairness and bias. Seth K warns that AI model's effectiveness can be compromised by low-quality data, including gaps [or] biases… Banks must prioritise robust data governance to maintain accurate, varied, and inclusive training data. In practice, teams counter this by rebalancing datasets, using synthetic data generation, and continuously monitoring model performance across subgroups.

Interpretability and explainability: Many cutting-edge AI models (deep learning, complex ensembles) operate as "black boxes" whose logic is opaque. Yet finance regulators often require explanations for decisions, especially if a transaction is declined or a customer is flagged. This makes interpretability a challenge. The industry is responding with **Explainable AI (XAI)** techniques. As noted, XAI tools can show *why* a transaction was flagged, revealing key features or rules that triggered it. For fraud teams, such transparency is crucial for auditing and justifying actions. However, building truly interpretable fraud models remains an open area of research.

Evolving fraud tactics: Fraudsters constantly innovate. The rise of online banking has led to new schemes (e.g. authorized push payments scams, supply chain invoicing fraud, social engineering). Moreover, emerging technologies pose new threats: for instance, *deepfake audio/video* could be used to spoof customer identity, prompting FinCEN alerts, and synthetic identities created by AI could fool KYC checks. In response, AI fraud systems must adapt quickly. Techniques like **anomaly detection** and continuous learning help – models can potentially spot an unusual pattern (e.g. a burst of pandemic-related phishing) even before rules exist for it. Indeed, generative AI is now being explored as a countermeasure: large language models (LLMs) could scan unstructured data (emails, support calls) to detect fraud signals, or even simulate attack scenarios. One report notes that next-generation AI – including LLMs – can *"detect fraud through anomaly pattern identification"*, further advancing analytics. However, ironically, the same generative tools might be used by fraudsters (e.g. generating realistic phishing messages), raising an "AI vs. AI" arms race.

Regulatory compliance and collaboration: Finally, aligning with policy is itself a challenge. As mentioned, privacy laws can impede data sharing. One promising trend to address this is federated learning: banks and fintechs collaboratively train a shared ML model without exchanging raw data. Seth K highlights federated learning as a way for institutions *to* collaborate on model training without sharing sensitive data, improving cross-border fraud detection. Similarly, cryptographic techniques and secure enclaves may allow pooling of insights while respecting privacy. (Seth, 2025) Other future directions include quantum computing (long-term potential to accelerate pattern recognition) and AI-powered blockchain analytics (e.g. flagging suspicious cryptocurrency transactions).

In summary, AI is transforming fraud prevention by making it more dynamic and data-driven, but it requires careful handling of biases, transparency, and compliance. As one industry expert puts it, financial institutions can turn AI-driven fraud detection into *"a competitive advantage"* by leveraging innovations to *"reduce false positives"* while bolstering security.

5. Conclusion

AI and machine learning have become central to fraud detection in digital payments. By applying supervised classifiers, unsupervised anomaly detectors, and real-time scoring, modern systems can sift through huge volumes of transaction data to flag illegitimate activity much faster than legacy rule-based methods. In both Europe and the U.S., fintech companies, neobanks, and traditional banks are deploying these tools: from startup challengers to global card networks. Early evidence suggests substantial gains (multi-billion-dollar fraud prevented by ML systems).

Regulators are catching up, though, as new AI regulations (such as the EU AI Act) and security requirements (such as PSD2 and AML laws) influence how these systems are developed and managed. Ensuring privacy, transparency, and equity still presents significant obstacles. These problems are being addressed by ongoing research and cooperation, such as in explainable AI, federated models, and regulatory sandboxes. Future developments like generative AI and quantum computing offer fraud fighters new resources, but they also bring new dangers. In general, everyone agrees that AI has a significant impact on enhancing payment security. When properly implemented, it increases fraud detection while lowering barriers for authorized users, bolstering the ecosystem's resilience and trust.

References

[1] Aros, L. H., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., & Rodríguez Barrero, M. S. (2024). Financial fraud detection through the application of machine learning techniques: a literature review. Humanities and Social Sciences Communications, 11, Article 1130. https://www.nature.com/articles/s41599-024-03606-0

[2] BaFin (Federal Financial Supervisory Authority). (n.d.). Big data, artificial intelligence and machine learning. Retrieved from https://www.bafin.de/EN/Aufsicht/FinTech/InnovativeFinanztechnologien/BDAI/BDAI_node_en.html

[3] Eurofi. (2024). AI Act: key measures and implications for financial services. Eurofi Regulatory Update (Sept 2024). https://www.eurofi.net/wp-content/uploads/2024/12/ii.2-ai-act-key-measures-and-implications-for-financial-services.pdf

[4] Seth K (2025). AI-Driven Payment Security: Transforming Fraud Detection and Safeguarding the Digital Economy https://www.analyticsinsight.net/artificial-intelligence/ai-driven-payment-security-transforming-fraud-detection-and-safeguarding-the-digital-economy

[5] Boehm, A. (2024, October 21). Analysis: Fraud challenges facing neobanks and fintechs. Finextra. https://www.finextra.com/blogposting/27060/analysis-fraud-challenges-facing-neobanks-and-fintechs

[6] Deutsche Bank. (2022, Sept 22). Deutsche Bank partners with Visa to prevent fraud in online retail [Press release]. https://www.db.com/news/detail/20220922-deutsche-bank-partners-with-visa-to-prevent-fraud-in-online-retail?language_id=1

[7] U.S. Department of the Treasury, Office of Financial Research. (2024). Uses, opportunities, and risks of artificial intelligence in financial services (Report). https://home.treasury.gov/system/files/136/Artificial-Intelligence-in-Financial-Services.pdf