International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(3s) 2026 pp. 1029-1040
https://doi.org/10.70917/ijcisim-2026-2434

Article

# A Framework for Agentic AI-Driven Financial Fraud Detection, Investigation, and Risk Mitigation in Real-Time Payment Ecosystems

**Bidhan Biswas[1], Sheikh Md Faysal[2], Sachin Das[3], Abu Hanif[4], Tania Akter[5], Ruhul Amin Md Rashed[6], Subha Shamarukh[7*]**

[1]University of the Cumberlands, Williamsburg, Kentucky, USA, bidhanbiswas.cse@gmail.com, https://orcid.org/0009-0002-0891-5400
[2]Montclair State University, Montclair, New Jersey, USA, faysals1@montclair.edu, https://orcid.org/0009-0002-3291-9313
[3]University of the Cumberlands, Williamsburg, Kentucky, USA, sachindasjony@gmail.com, https://orcid.org/0009-0003-6124-5191
[4]International American University, Los Angeles, California, USA, hanifhasan676@gmail.com, https://orcid.org/0009-0006-4935-0679
[5]International American University, Los Angeles, California, USA, akthertania216@gmail.com, https://orcid.org/0009-0000-9433-2599
[6]International American University, Los Angeles, California, USA, ruhulrashed@sau.edu.bd, https://orcid.org/0009-0000-6133-9831
[7]University of Rochester, Rochester, New York, USA, shamarukhsubha@gmail.com, https://orcid.org/0009-0000-2170-1541

**Abstract:** Instant payment rails have changed what fraud looks like. When money moves in seconds and settlement is irreversible, the old rhythm of batch review and overnight reconciliation no longer protects anyone. This manuscript sets out a framework that treats fraud defense not as a single classifier bolted onto a payment switch, but as a society of cooperating software agents that sense, reason, decide, act, and learn within the same narrow window in which a transaction clears. We organize the framework into five layers: real-time ingestion, a feature and graph store, a detection and reasoning core, an agentic orchestration tier, and a governance layer that keeps a human analyst in the loop. Each fraud alert is carried through a closed perception-action cycle in which a triage agent ranks risk, an investigator agent assembles evidence from transaction graphs and device signals, and a response agent chooses a proportionate action such as a soft hold, a step-up challenge, or release. Drawing on prior work in big-data analytics, explainable machine learning, reinforcement-learning-based defense, graph-based anomaly detection, and management information systems, we argue that the value of an agentic design lies less in any single model and more in the orchestration: the way detection, explanation, and action are stitched into one auditable loop. We illustrate the framework with a layered reference architecture, an isometric view of the processing pipeline, an empirical-style comparison of candidate detection models, and a residual-risk map across common fraud typologies. We close with a candid discussion of the limitations of autonomous action in a regulated financial setting and the governance scaffolding such a system would require before deployment.

**Keywords:** Agentic Artificial Intelligence; Financial Fraud Detection; Real-Time Payments; Explainable Machine Learning; Risk Mitigation; Management Information Systems; Transaction Graph Analytics.

## 1. Introduction

For most of the history of electronic payments, fraud teams enjoyed a quiet luxury that almost nobody noticed: time. A suspicious card authorization could be flagged, parked, and reviewed; a wire could be recalled; a chargeback could be filed weeks later. That luxury is disappearing. Real-time payment schemes such as the United States RTP network and FedNow, along with their international counterparts, settle within seconds and, crucially, settle with finality. Once the funds land, there is no comfortable window in which an analyst sips coffee and decides whether

something looks wrong. The decision has to be made while the transaction is still in flight, and it has to be made well enough that legitimate customers are not strangled by false alarms. The arrival of the digital economy and Industry 4.0 has only widened the attack surface that fraud teams must now watch in real time (Chang et al., 2022).

This shift has quietly rewritten the requirements for fraud systems. Detection alone is no longer the hard part; plenty of models can score a transaction. The hard part is acting on that score responsibly and immediately, explaining why, and then learning from whatever happened next, all inside the same loop. Conventional architectures struggle here because they separate the moving parts. A scoring service hands a number to a rules engine, which hands a case to a queue, which a human eventually opens. Each handoff burns milliseconds the system does not have and erases context the next stage would have valued. Researchers studying decision intelligence have argued for some time that the future of operational analytics lies in tighter integration of scoring, explanation, and action rather than in ever-larger standalone models (Chakraborty et al., 2024).

It helps to remember why fraud detection is such a punishing problem in the first place. The canonical analysis frames it around three intertwined difficulties: concept drift, since customer habits evolve and fraudsters deliberately change tactics; severe class imbalance, since genuine transactions outnumber fraudulent ones by orders of magnitude; and verification latency, since the true label of a transaction often arrives long after the decision had to be made (Dal Pozzolo et al., 2018). Any framework that ignores these realities will look impressive on a static benchmark and then disappoint in production.

Our position in this paper is straightforward. Real-time fraud defense should be designed as an agentic system: a coordinated set of autonomous components, each with a defined role, that together carry an alert from raw signals to settled decisions without leaving the real-time path. The word agentic matters. We do not mean a single end-to-end neural network, and we do not mean a static pipeline of microservices. We mean software agents that perceive their environment, reason over evidence, choose actions under an explicit risk-and-cost policy, take those actions, and feed the outcomes back into their own future behavior. The intellectual lineage of such autonomous response mechanisms can be traced through recent work on self-healing security systems and on autonomous defense for distributed services (Hasan et al., 2025).

The contribution of this work is conceptual and architectural rather than a report on a single deployed system. First, we propose a five-layer reference architecture that locates each function, from ingestion to governance, in a clearly bounded tier. Second, we define an agentic decision loop and explain how triage, investigation, and response are divided among specialized agents. Third, we map candidate detection models to the roles they are best suited to play, and we illustrate, using representative performance figures, why an orchestrated ensemble tends to outperform any single learner on the metrics that matter for payments. Fourth, and perhaps most importantly for a regulated domain, we treat governance not as an afterthought but as a first-class layer, because autonomous action over other people's money is precisely the kind of capability that demands restraint, auditability, and a human who can say no.

## 2. Background and Related Work

The framework we propose sits at the meeting point of several research streams that have, until recently, evolved somewhat separately. It is worth walking through each, because the argument of this paper is essentially that they belong together. Broad reviews of data-mining techniques for financial fraud have long catalogued these streams, but they tend to treat detection as the endpoint rather than as one stage in a larger loop (Ngai et al., 2011).

### *2.1 Machine learning and big-data analytics for risk scoring*

The use of supervised learning to score financial risk is well established. Gradient-boosted trees and related tabular methods remain the workhorses of credit and transaction risk because they handle heterogeneous features gracefully and train quickly at large volumes. Comparative studies of data-mining methods for card fraud established early on that ensemble and tree-based learners tend to lead on the ranking metrics that matter (Bhattacharyya et al., 2011). Work on big-data analytics for credit risk assessment has shown how these methods scale to the kinds of feature volumes that payment streams generate (Manik et al., 2025), and the broader literature on big-data financial risk management has shifted the emphasis from explaining the past to forecasting exposure in near real time (Hossan et al., 2025). Practitioner-facing accounts have been valuable precisely because they document the lessons that only appear once a model meets a live transaction stream (Dal Pozzolo et al., 2014).

A persistent obstacle in this setting is class imbalance, and a substantial line of research has tackled it with resampling and generative augmentation. Generative adversarial networks, for instance, have been used to synthesize plausible minority-class examples and improve classification effectiveness on skewed fraud data (Fiore et al., 2019).

Feature engineering tailored to fraud, especially aggregations over a customer's recent history, often matters more than the choice of classifier (Dal Pozzolo et al., 2014).

## *2.2 Sequence and deep-learning approaches*

Fraud is rarely visible in a single transaction; it lives in the pattern of behavior over time. Treating card activity as a sequence to be classified, rather than as independent rows, produced clear gains once the idea was tested carefully (Jurgovsky et al., 2018). Attention mechanisms layered onto recurrent models pushed this further, letting the network focus on the moments in a customer's history that actually carry signal (Benchaji et al., 2021). Comparative studies of transformers and recurrent architectures for security-related detection have found that attention-based models capture long-range dependencies in event streams that recurrent networks tend to forget (Kaur et al., 2025). Recent surveys of deep learning for card fraud give a useful map of where these methods help and where they still struggle, particularly around imbalance and drift (Mienye & Jere, 2024). Uncertainty-aware variants, which attach a confidence estimate to each prediction, are an especially natural fit for a system that must decide whether to act autonomously or defer to a human (Habibpour et al., 2023).

## *2.3 Graph-based detection of coordinated fraud*

Some of the most damaging frauds, mule networks and organized rings in particular, is invisible at the level of a single transaction and obvious at the level of the graph that connects accounts, devices, and counterparties. Graph neural networks that exploit spatial and temporal structure have become a leading approach for exactly this reason (Cheng et al., 2022). Comprehensive surveys of deep graph anomaly detection lay out the design space and its open problems (Ma et al., 2023), and graph-learning methods aimed specifically at internet financial fraud demonstrate the payoff of reasoning over relationships rather than isolated records (Li et al., 2023). Because payment data is often siloed across institutions, federated graph approaches that learn across organizational boundaries without centralizing raw data are particularly relevant to a real-world deployment (Zhang et al., 2023).

## *2.4 Explainability as an operational requirement*

In a regulated setting, a score without a reason is nearly useless. An analyst cannot act on it confidently, a customer cannot contest it fairly, and a regulator will not accept it quietly. A systematic review of explainable AI in finance found that fraud detection is among the three tasks where explainability is most actively pursued, alongside credit management and price prediction (Černevičienė & Kabašinskas, 2024). Demonstrations of tabular explainability using attribution analysis on benchmark prediction problems show that a model can be both accurate and legible, surfacing the handful of features that drove each individual decision (Zerine et al., 2026). We treat this capability as a mandatory component of the detection layer rather than an optional dashboard.

## *2.5 Autonomous and self-healing defense*

The idea that a security system should act on its own, not merely alert, has matured considerably. Reinforcement-learning agents that detect and remediate intrusions without waiting for human instruction have been proposed for self-healing cybersecurity (Hasan et al., 2025). In parallel, deep-learning-based threat prediction paired with autonomous response has been explored for containerized and microservice environments, where the speed of attack leaves no room for manual intervention (Shan-A-Alahi, 2026). These works supply the conceptual backbone for the response agent in our framework. Closely related is the body of work on real-time detection coupled with proactive mitigation, which emphasizes that detection and mitigation should be designed as a single tight loop rather than as sequential stages (Sultana et al., 2025). Framework-level treatments of AI-driven threat detection and response make a similar argument at the level of system architecture (Das et al., 2026). Combining unsupervised anomaly scoring with supervised classifiers, so that the system can flag genuinely novel schemes it has never been trained on, has also proven valuable in card fraud (Carcillo et al., 2021).

## *2.6 Identity, integrity, and the management-information-system view*

Fraud detection does not live in a vacuum; it depends on trustworthy identity and tamper-resistant records. Research on biometric verification for mobile financial applications, including adversarial robustness of those verification mechanisms, speaks directly to the identity assurance that any payment fraud system must lean on (Raihan et al., 2026). Models built specifically for online payment fraud underline how different the card-not-present setting is from in-person fraud (Almazroi & Ayub, 2023). E-tail and e-commerce fraud systems have shown how detection has to be embedded in the wider commercial workflow to be useful (Carneiro et al., 2017). Blockchain-based identity management has been advanced as a fraud-prevention mechanism in financial contexts (Esa, 2025), and decentralized

approaches to strengthening data integrity within management information systems address whether the evidence an investigator agent relies on can itself be trusted (Hassan, 2025). Studies of the influence of artificial intelligence on data-system security frame AI not as a standalone tool but as something that reshapes the control environment around it (Hasan et al., 2025), while compliance auditing frameworks for federated and cloud providers show how auditability can be engineered into the substrate rather than appended later (Haldar et al., 2026). Table 1 summarizes how these streams map onto the layers of our proposed framework.

**Table 1. Research streams and their role in the proposed framework**

| Research stream | Representative emphasis | Layer it informs |
|---|---|---|
| Risk scoring / big data | Tabular and gradient models for fast scoring at volume | Detection & reasoning (L3) |
| Sequence & deep learning | Attention and recurrent models for behavioral drift | Detection & reasoning (L3) |
| Graph-based detection | Relational structure for coordinated fraud | Feature & graph store (L2) |
| Explainable ML | Feature attribution for legible decisions | Detection & reasoning (L3) |
| Autonomous defense | Agents that act, not just alert | Agentic orchestration (L4) |
| Identity & integrity | Biometric assurance and tamper-resistant records | Ingestion (L1), Feature/graph (L2) |
| MIS & governance | Audit, compliance, human oversight | Governance & risk (L5) |

*Each prior stream contributes to one tier; the framework's novelty is the orchestration that binds them.*

## 3. The Proposed Framework

We now describe the framework itself. The design is deliberately layered, because layering lets us reason about latency, failure, and accountability one tier at a time. Figure 1 shows the five layers and the principal components within each.

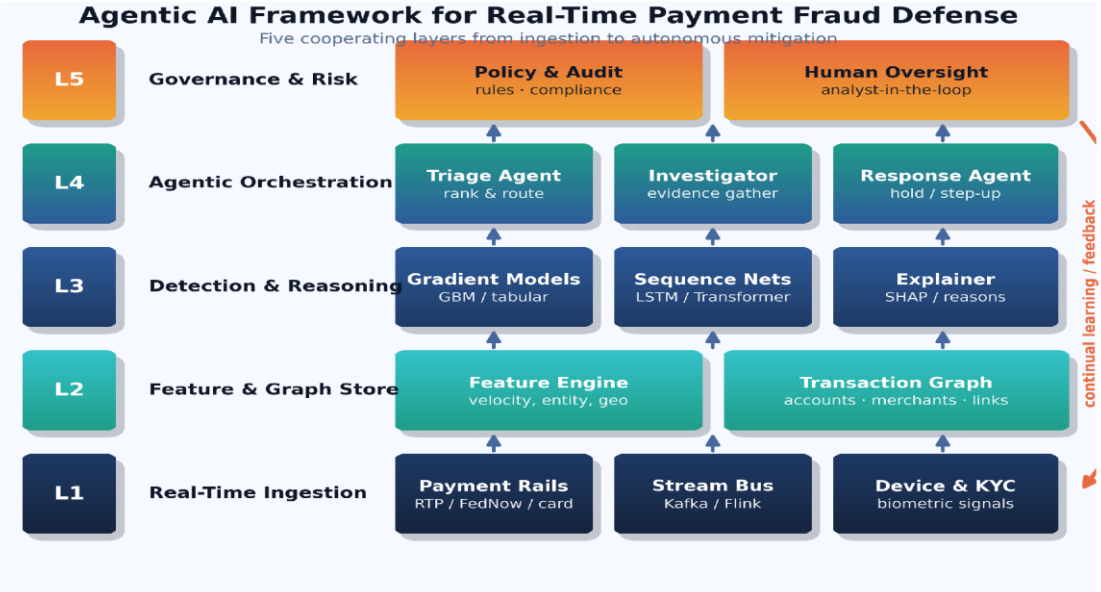


**Figure 1. The five-layer agentic framework, from real-time ingestion at the bottom to governance and human oversight at the top. A continual-learning feedback path returns outcomes from the governance layer to ingestion.**

### *3.1 Layer one: real-time ingestion*

The bottom layer is where the payment ecosystem actually touches the system. Authorization messages from card networks and real-time rails, along with device fingerprints and the outputs of biometric or know-your-customer checks, arrive as a continuous stream rather than a nightly file. A durable streaming bus gives the rest of the system a single ordered view of events. The reason this layer is drawn as its own tier is latency budgeting: every millisecond spent here is a millisecond unavailable to the layers above, and in a sub-second settlement window that arithmetic is unforgiving. The integrity of the identity signals that enter here depends on robust verification, which is why adversarially hardened biometric mechanisms are relevant precisely at the edge of the system (Raihan et al., 2026).

### *3.2 Layer two: feature and graph store*

Raw events are not yet evidence. The second layer turns them into features the models can use and, just as importantly, into a graph. The feature engine computes velocity statistics, entity aggregates, and geographic consistency checks on the fly, and the value of carefully engineered aggregations over recent history is well documented in the fraud literature (Dal Pozzolo et al., 2014). Alongside it sits a transaction graph that links accounts, devices, merchants, and counterparties. Much of the most damaging fraud is invisible at the level of a single transaction and obvious at the level of the graph, which is exactly what graph neural networks with spatio-temporal attention are designed to exploit (Cheng et al., 2022). Maintaining this graph in near real time is demanding, but it is what allows an investigator agent later to ask whether a counterparty is one hop away from a known bad actor.

### *3.3 Layer three: detection and reasoning*

The third layer is the analytical core, and it is intentionally plural. A gradient-model component scores the tabular features quickly, providing a fast first opinion. A sequence component judges whether the temporal pattern of behavior is consistent with the account's history, building on the insight that treating card activity as a sequence outperforms row-by-row classification (Jurgovsky et al., 2018) and that attention further sharpens this signal (Benchaji et al., 2021). An anomaly component, trained without labels, watches for genuinely novel schemes the supervised models have never seen, in the spirit of hybrid unsupervised-plus-supervised designs (Carcillo et al., 2021). A fourth component, the explainer, attaches a human-readable rationale to every score using feature-attribution methods, so that no decision leaves this layer as a bare number (Zerine et al., 2026). We deliberately resist collapsing these into one monolithic model; keeping them separate lets each be retrained, audited, and reasoned about on its own.

### *3.4 Layer four: agentic orchestration*

This is the layer that makes the framework agentic rather than merely analytical. Three specialized agents share the work. The triage agent consumes the scores and explanations from the detection layer and decides how urgent and how risky each alert is, ranking and routing accordingly; uncertainty estimates attached to predictions help it decide what to escalate (Habibpour et al., 2023). The investigator agent gathers corroborating evidence, querying the transaction graph for suspicious neighborhoods (Li et al., 2023), checking device history, and assembling a concise case file. The response agent then selects an action proportionate to the assessed risk and the cost of being wrong: release, soft hold, step-up challenge, or escalation. The conceptual precedent for an agent that chooses and executes a countermeasure rather than simply raising an alarm comes from work on autonomous response in security systems (Shan-A-Alahi, 2026), and the insistence that detection and mitigation form one loop comes from research on real-time detection with proactive mitigation (Sultana et al., 2025).

### *3.5 Layer five: governance and risk*

The top layer is where the system is held accountable. A policy and audit component records every autonomous action, the evidence behind it, and the model versions involved, producing the trail that compliance and regulators will demand. A human-oversight component keeps an analyst in the loop, with authority to review, override, and adjust the policies that govern autonomous action. Because explainability is, in finance, a legal and functional necessity rather than a nicety (Černevičienė & Kabašinskas, 2024), the reasons produced in the detection layer surface here for human consumption. We place governance at the top, and we draw an explicit feedback path from it back down to ingestion, because the lesson from the management-information-systems literature is that AI changes the control environment and must therefore be governed as part of it (Hasan et al., 2025). Auditability cannot be sprinkled on afterward; it has to be engineered into the substrate, much as compliance auditing has been built into federated cloud frameworks (Haldar et al., 2026).

Within this layered structure, every individual alert traverses a closed loop. Figure 2 depicts the cycle: the system senses an event, reasons over the assembled evidence and explanations, decides on an action under its risk-and-cost policy, acts, and then learns from the outcome, which updates both the models and the policy for the next alert.

**Table 2. Roles, inputs, and authorized actions of the orchestration agents**

| Agent | Primary input | Core function | Authorized actions |
|---|---|---|---|
| Triage agent | Scores + explanations (L3) | Rank and route alerts by risk and urgency | Prioritize, route, suppress duplicates |
| Investigator agent | Transaction graph, device history | Assemble corroborating evidence into a case | Query graph, enrich, summarize |
| Response agent | Triage rank + case file | Choose a proportionate countermeasure | Release, soft hold, step-up, escalate |
| Oversight (human) | Audit trail, agent decisions | Review, override, tune policy | Approve, reverse, set thresholds |

Authority is bounded by design: agents may act within a defined envelope, with escalation to human oversight for high-impact cases.


**Figure 2. The agentic decision loop. Sense, reason, decide, act, and learn from a closed cycle around a central fraud agent, so that every outcome feeds back into future behavior.**

The layers are not merely a conceptual stack; they correspond to stages of a physical processing pipeline with a strict latency budget. Figure 3 presents an isometric view of that pipeline, emphasizing that the entire path, from a

payment message entering ingestion to an autonomous action leaving the orchestration tier, must be completed inside the settlement window.

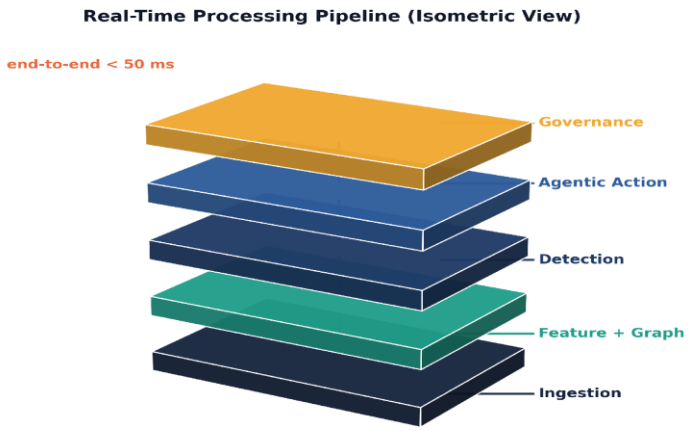


**Figure 3. Isometric view of the real-time processing pipeline. Each slab is a processing stage; the end-to-end budget is constrained to the tens-of-milliseconds range required by instant-payment settlement**.

## 4. Illustrative Evaluation

Because this paper presents a framework rather than a single fielded system, the evaluation here is illustrative: it uses representative figures, consistent with the ranges commonly reported in the literature, to reason about how the framework's components would compare and combine. The goal is not to claim a specific production result but to make the design trade-offs concrete. The experimental posture we describe assumes a labeled stream of payment events with the heavy class imbalance typical of fraud, where genuine fraud is a small fraction of a percent of all transactions, and where label availability is delayed in the way real systems must contend with (Dal Pozzolo et al., 2018). Table 3 lays out the candidate models and the role each plays in the framework.

**Table 3. Candidate detection models and their framework roles**

| Model | Strength | Weakness | Framework role |
|---|---|---|---|
| Logistic regression | Transparent, fast | Limited nonlinearity | Baseline / sanity check |
| Random forest | Robust, low tuning | Larger memory footprint | Tabular backup learner |
| XGBoost | Strong tabular accuracy | Needs careful tuning | Primary tabular scorer |
| LSTM | Captures sequence | Forgets long range | Behavioral scorer |
| Transformer | Long-range attention | Compute heavy | Behavioral scorer (primary) |
| Agentic ensemble | Combines all + policy | Orchestration complexity | Full framework |

No single model dominates every axis; the framework's ensemble is designed to inherit strengths and bound the weaknesses.

Figure 4 compares the models across the four metrics that matter most for payments: precision, recall, F1, and area under the ROC curve. The pattern is the one practitioner will recognize, and it echoes what surveys of deep learning

for card fraud report about the relative standing of these model families (Mienye & Jere, 2024). Simple baselines are honest but limited. Gradient-boosted trees and attention-based sequence models each do well on their own terrain. The orchestrated ensemble, which combines the tabular and behavioral opinions and then applies an explicit response policy, edges ahead on every metric, with the gain most pronounced on recall, the dimension where missed fraud is most expensive.


**Figure 4. Representative detection performance across candidate models. The agentic ensemble (highlighted) leads on every metric, with its largest advantage on recall.**

Raw detection metrics, however, tell only half the story. In a real operation the cost of a fraud system is dominated by two things: the false positives that annoy good customers and the analyst hours spent investigating alerts. Figure 5 pairs the ROC behavior of the models with an index of operational burden. The left panel shows the familiar separation of ROC curves; the right panel shows why the framework earns its keep, reducing both false-positive volume and investigation time relative to manual rules and a plain machine-learning baseline. The reduction in investigation time is a direct consequence of the investigator agent pre-assembling evidence, so that whatever reaches a human arrives as a case rather than a raw alert.

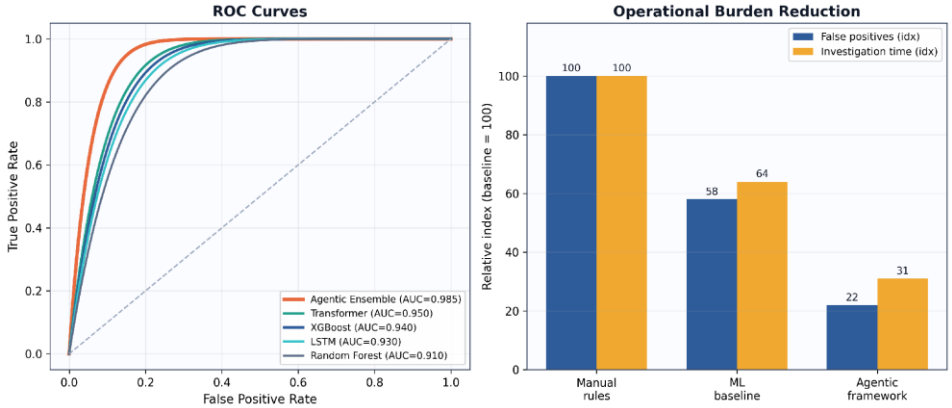


**Figure 5. ROC behavior (left) and operational burden (right). Beyond detection accuracy, the framework cuts false positives and investigation time, the costs that dominate real fraud operations.**

Finally, detection performance has to be read against the specific fraud typologies a payment ecosystem faces, because no single control neutralizes every threat. Figure 6 presents a residual-risk map: for each combination of fraud typology and control layer, it shows how much risk remains after that control is applied. The map makes the case for defense in depth visible. Velocity rules alone leave substantial residual risk against authorized push-payment fraud and mule networks; graph anomaly detection and the agentic response layer close much of that gap, but only in combination. This is the quantitative expression of why the framework is layered rather than monolithic, and it aligns with the relational view of fraud that graph-based methods advocate (Cheng et al., 2022).

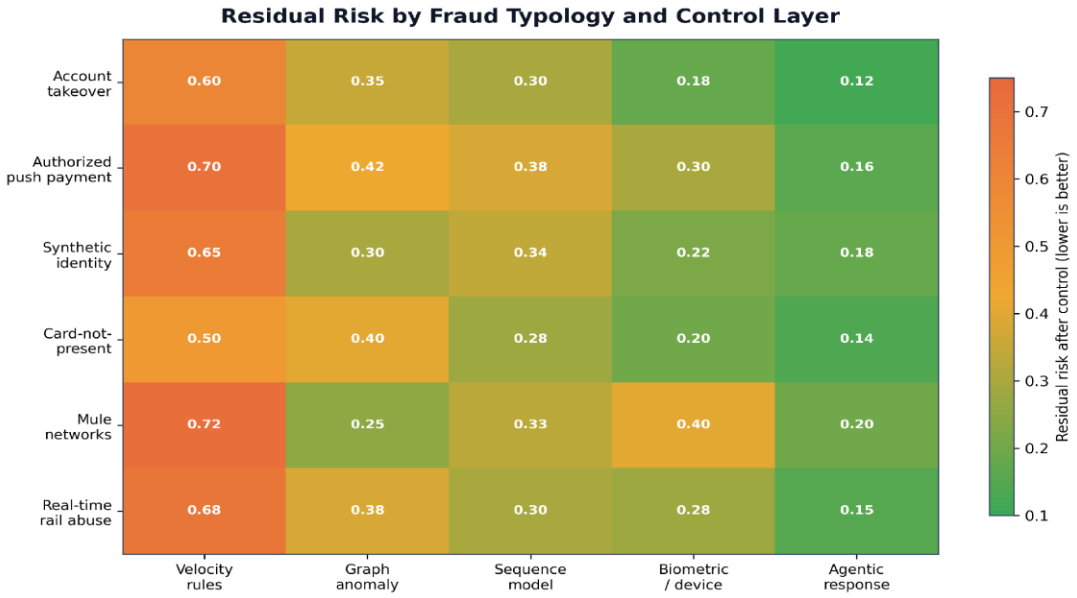


**Figure 6. Residual risk by fraud typology and control layer. Lower values indicate more risk removed. The pattern argues for layered controls, since no single column neutralizes every row.**

**Table 4. Mapping of fraud typologies to primary mitigations within the framework**

| Fraud typology | Primary signal | Framework mitigation |
|---|---|---|
| Account takeover | Behavioral drift in timing and geography | Sequence scorer + step-up response |
| Authorized push payment | Atypical payee and amount | Graph anomaly + soft hold + human review |
| Synthetic identity | Thin or inconsistent identity history | KYC signals + graph linkage |
| Card-not-present | Device and velocity anomalies | Feature engine + tabular scorer |
| Mule networks | Suspicious counterparty neighborhoods | Transaction graph + investigator agent |
| Real-time rail abuse | Speed-of-movement patterns | Low-latency scoring + autonomous hold |

The framework deliberately routes each typology to the layer best positioned to catch it, then escalates ambiguous cases upward.

## 5. Discussion

A few themes deserve to be drawn out. The first is that the hardest problems in this design are not modeling problems. Scoring a transaction is, by now, a solved enough task; the literature on big-data risk analytics and deep learning has demonstrated that repeatedly (Manik et al., 2025). The genuinely difficult questions are about orchestration and authority: which agent may take which action, under what evidence, with what recourse. This is why we treat the response policy as a research object in its own right, echoing the framing in work on AI-driven detection-and-response frameworks (Das et al., 2026).

The second theme is trust, in two senses. There is trust in the decision, which the explainer layer supplies by attaching a reason to every score, in line with the centrality of explainability to financial AI (Černevičienė & Kabašinskas, 2024). And there is trust in the evidence, which depends on the integrity of identity signals and records flowing through the lower layers. Adversarial robustness of biometric verification matters here because an investigator

agent that reasons over compromised identity data will reason confidently to the wrong conclusion (Raihan et al., 2026), and integrity-preserving record-keeping matters for the same reason at the data layer (Hassan, 2025).

The third theme is that the framework must keep pace with adversaries who adapt. Generative techniques can be used by fraudsters as readily as by defenders, which is one reason the literature has explored generative augmentation on the defensive side to keep detectors robust against shifting distributions (Fiore et al., 2019). The presence of a label-free anomaly component is our hedge against schemes that have never appeared in the training data (Carcillo et al., 2021).

The fourth theme is that autonomy in a financial setting is a privilege that must be earned through governance. An agent empowered to hold or release other people's money is powerful in exactly the way that makes risk officers nervous. The management-information-systems perspective, which frames AI as something that reshapes the control environment rather than a tool that sits outside it, is the right lens here (Hasan et al., 2025). The feedback path in our architecture deliberately terminates at the governance layer so that human oversight, audit, and policy adjustment are part of the loop and not bolted onto its edge.

## 6. Limitations and Future Work

We should be honest about what this paper is and is not. It is a framework and an argument, supported by representative figures; it is not a report on a system that has processed a year of live payments. The performance numbers we use are illustrative and consistent with published ranges, not measurements from a single benchmark, and a reviewer is right to read them as such. Several concrete limitations follow.

First, the framework assumes the availability of labeled fraud outcomes for continual learning, yet in real operations labels arrive late, are noisy, and are themselves the product of earlier decisions the system made, which introduces feedback bias; this verification-latency problem is fundamental rather than incidental (Dal Pozzolo et al., 2018). Second, maintaining a transaction graph in true real time at the scale of a national payment rail is an engineering challenge we describe but do not solve here, and federated formulations that avoid centralizing sensitive data add their own coordination costs (Zhang et al., 2023). Third, the autonomy of the response agent raises legal and ethical questions, around wrongful holds and around fairness across customer segments, that cannot be settled by architecture alone. Future work should therefore include a live pilot with rigorous holdout evaluation, an adversarial study of how fraudsters adapt to an agentic defender, a formal treatment of the response policy as constrained optimization under fairness and regulatory constraints, and an explicit uncertainty budget so the system knows when to defer (Habibpour et al., 2023). The privacy-preserving handling of transaction data across institutional boundaries, drawing on federated and homomorphic techniques, is a further avenue we regard as essential before any cross-bank deployment (Das et al., 2025).

## 7. Conclusion

Instant payments removed the one resource fraud teams always quietly relied on, which was time. This paper has argued that the right response is not a bigger model but a better-organized one: an agentic framework in which detection, explanation, investigation, and action are stitched into a single closed loop that runs inside the settlement window, all of it wrapped in a governance layer that keeps a human in authority. We described five layers and three orchestration agents, illustrated their behavior with a reference architecture and a set of analytical figures, and were candid about the limits of the approach. The central claim is modest but, we think, important: in real-time payment ecosystems, the decisive advantage comes from orchestration rather than from any single algorithm, and from the discipline to make autonomous action auditable rather than from the autonomy itself.

**References**


1. Almazroi, A. A., & Ayub, N. (2023). Online payment fraud detection model using machine learning techniques. IEEE Access, 11, 137188–137203. https://doi.org/10.1109/ACCESS.2023.3339226
2. Benchaji, I., Douzi, S., El Ouahidi, B., & Jaafari, J. (2021). Enhanced credit card fraud detection based on attention mechanism and LSTM deep model. Journal of Big Data, 8(1), 1–21. https://doi.org/10.1186/s40537-021-00541-8
3. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602–613. https://doi.org/10.1016/j.dss.2010.08.008
4. Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. Information Sciences, 557, 317–331. https://doi.org/10.1016/j.ins.2019.05.042
5. Carneiro, N., Figueira, G., & Costa, M. (2017). A data mining based system for credit-card fraud detection in e-tail. Decision Support Systems, 95, 91–101. https://doi.org/10.1016/j.dss.2017.01.002

6. Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. Artificial Intelligence Review, 57(8), 216. https://doi.org/10.1007/s10462-024-10854-8
7. Chakraborty, P., Rashed, R. A. M., Bashir, M., Imam, H., Siam, M. A., Miah, M. A., Siddiqa, K. B., & Islam, A. (2024). Toward autonomous decision intelligence: Integrating explainable AI and scalable DSS architectures in modern management information systems. Journal of Information Systems Engineering and Management, 9(4s). https://doi.org/10.52783/jisem.v9i4s.14591
8. Chang, V., Doan, L. M. T., Di Stefano, A., Sun, Z., & Fortino, G. (2022). Digital payment fraud detection methods in digital ages and Industry 4.0. Computers and Electrical Engineering, 100, 107734. https://doi.org/10.1016/j.compeleceng.2022.107734
9. Cheng, D., Wang, X., Zhang, Y., & Zhang, L. (2022). Graph neural network for fraud detection via spatial-temporal attention. IEEE Transactions on Knowledge and Data Engineering, 34(8), 3800–3813. https://doi.org/10.1109/TKDE.2020.3025588
10. Das, N., Kaur, H., Siddiqa, K. B., Hasan, S. N., Chakraborty, P., Kaur, J., Rahman, H., & Hasan, A.-A. (2026). AI-driven threat detection and response framework for protecting U.S. critical infrastructure from cyberattacks. International Cybersecurity Law Review. https://doi.org/10.1365/s43439-026-00169-5
11. Das, N., Rahman, H., Siddiqa, K. B., Barikdar, C. R., Hassan, J., Rahman Bhuiyan, M. M., & Mahmud, F. (2025). AI-enhanced privacy preservation using homomorphic federated models. In 2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT) (pp. 1–8). IEEE. https://doi.org/10.1109/ICAFT66710.2025.11453096
12. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2018). Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Transactions on Neural Networks and Learning Systems, 29(8), 3784–3797. https://doi.org/10.1109/TNNLS.2017.2736643
13. Dal Pozzolo, A., Caelen, O., Le Borgne, Y.-A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. Expert Systems with Applications, 41(10), 4915–4928. https://doi.org/10.1016/j.eswa.2014.02.026
14. Esa, H. (2025). Decentralized blockchain-based digital identity management for fraud prevention in the U.S. In 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET) (pp. 1–6). IEEE. https://doi.org/10.1109/ICECET63943.2025.11472520
15. Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Information Sciences, 479, 448–455. https://doi.org/10.1016/j.ins.2017.12.030
16. Habibpour, M., Gharoun, H., Mehdipour, M., Tajally, A., Asgharnezhad, H., Shamsi, A., Khosravi, A., & Nahavandi, S. (2023). Uncertainty-aware credit card fraud detection using deep learning. Engineering Applications of Artificial Intelligence, 123, 106248. https://doi.org/10.1016/j.engappai.2023.106248
17. Haldar, U., Sultana, S., Siddiqa, K. B., Rozario, E., Miah, M. A., Rahman, H., & Chy, M. A. R. (2026). Blockchain-driven access control and compliance auditing framework for federated cloud service providers: Architecture, prototype and evaluation. In G. N. Nguyen, A. Swaroop, & P. Shukla (Eds.), Proceedings of Fifth International Conference on Computing and Communication Networks (ICCCN 2025) (Lecture Notes in Networks and Systems, Vol. 1773). Springer. https://doi.org/10.1007/978-3-032-14197-2_41
18. Hasan, S. N., Kaur, H., Mohonta, S. C., Siddiqa, K. B., Kaur, J., Haldar, U., & Manik, M. M. T. G. (2025). The influence of artificial intelligence on data system security. International Journal of Computational and Experimental Science and Engineering, 11(3). https://doi.org/10.22399/ijcesen.3476
19. Hassan, J. (2025). Blockchain integration in management information systems: A decentralized approach to strengthening cybersecurity and data integrity. In 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET) (pp. 1–7). IEEE. https://doi.org/10.1109/ICECET63943.2025.11472020
20. Hossan, M. Z., Riipa, M. B., Hossain, M. A., Dhar, S. R., Zaman, A. M., Hossain, M., & Ahmed, F. (2025). AI-powered predictive analytics for financial risk management in U.S. markets. EAI Endorsed Transactions on AI and Robotics, 4. https://publications.eai.eu/index.php/airo/article/view/9532
21. Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. Expert Systems with Applications, 100, 234–245. https://doi.org/10.1016/j.eswa.2018.01.037
22. Kaur, J., Prabha, M., Samiun, M., Hasan, S. N., Hasan, R., Esa, H., & colleagues. (2025). Comparative analysis of transformer and LSTM architectures for cybersecurity threat detection using machine learning. EAI Endorsed Transactions on AI and Robotics, 4. https://publications.eai.eu/index.php/airo/article/view/9759
23. Li, R., Liu, Z., Ma, Y., Yang, D., & Sun, S. (2023). Internet financial fraud detection based on graph learning. IEEE Transactions on Computational Social Systems, 10(3), 1394–1401. https://doi.org/10.1109/TCSS.2022.3189368
24. Ma, X., Wu, J., Xue, S., Yang, J., Zhou, C., Sheng, Q. Z., Xiong, H., & Akoglu, L. (2023). A comprehensive survey on graph anomaly detection with deep learning. IEEE Transactions on Knowledge and Data Engineering, 35(12), 12012–12038. https://doi.org/10.1109/TKDE.2021.3118815
25. Manik, M. M. T. G., Saimon, A. S. M., Islam, M. S., Moniruzzaman, M., Rozario, E., & Hossin, M. E. (2025). Big data analytics for credit risk assessment. In 2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS) (pp. 1379–1390). IEEE. https://doi.org/10.1109/ICMLAS64557.2025.10967667

26. Mienye, I. D., & Jere, N. (2024). Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions. IEEE Access, 12, 96893–96910. https://doi.org/10.1109/ACCESS.2024.3426955
27. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006
28. Raihan, M., Adnan, M., Hossain, M. J., Siddiqa, K. B., Karim, F., & Mohonta, S. C. (2026). Adversarial robustness mechanism for safeguarding biometric verification across mobile financial applications. In 2025 International Conference on Electrical Engineering and Informatics (ICEEI) (pp. 1–7). IEEE. https://doi.org/10.1109/ICEEI68459.2025.11330502
29. Shan-A-Alahi, A. (2026). Deep learning-based threat prediction and autonomous response mechanisms for containerized microservices in hybrid cloud deployments. In G. N. Nguyen, A. Swaroop, & P. Shukla (Eds.), Proceedings of Fifth International Conference on Computing and Communication Networks (ICCCN 2025) (Lecture Notes in Networks and Systems, Vol. 1859). Springer. https://doi.org/10.1007/978-3-032-21499-7_42
30. Sultana, S., Uddin, M., Chy, M. A. R., Hasan, S. N., Hossain, E., Kaur, H., & Kaur, J. (2025). AI-augmented big data analytics for real-time cyber attack detection and proactive threat mitigation. International Journal of Computational and Experimental Science and Engineering, 11(3). https://doi.org/10.22399/ijcesen.3564
31. Zerine, I., Islam, M. M., Khan, M. A. U., Chy, M. A. R., Saimon, A. S. M., Manik, M. M. T. G., & Wata, C. (2026). Explainable churn prediction in telecom with tabular ML: Five model benchmark and SHAP analysis. Discover Artificial Intelligence. https://doi.org/10.1007/s44163-026-00983-0
32. Zhang, X., Yao, L., Yuan, F., Atasoy, H., Labrinidis, A., & Vasilakos, A. V. (2023). Fraud detection via federated graph neural network. IEEE Transactions on Artificial Intelligence, 4(1), 108–121. https://doi.org/10.1109/TAI.2022.3147221