Reinforcement Learning for Adaptive Fraud Detection in Changing Financial Environments

Gaetano Lachmann
School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
gaetanowork@uc.edu

Sante Luzi
Department of Computer Science, George Mason University, Fairfax, VA, USA.
sante1979@gmu.edu

Ligia Lorenzo
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
ligiamail@unr.edu

Lamberto Bongiorno
Department of Computer Science, University of New Hampshire, Durham, NH, USA.
lambertobongiorno@unh.edu

Abstract

Financial fraud detection is increasingly confronted with environments that change faster than conventional supervised models can be retrained, governed, and operationally trusted. This paper examines reinforcement learning as an adaptive decision architecture for fraud detection in shifting financial ecosystems. Rather than treating fraud detection as a static classification task, the analysis positions it as a sequential socio-technical decision problem involving exploration, delayed feedback, adversarial adaptation, and institutional accountability. The paper discusses how reinforcement learning can support dynamic thresholding, transaction sequencing, alert triage, and multi-agent coordination across payment networks, banks, merchants, and regulators. It also analyzes structural trade-offs involving false positives, customer friction, enforcement capacity, explainability, fairness, and operational resilience. Particular attention is given to data infrastructure, streaming pipelines, feature governance, model monitoring, and deployment architectures that must sustain adaptive behavior without introducing uncontrolled instability. The paper further considers policy implications, including auditability, model risk management, consumer protection, and cross-border regulatory alignment. It argues that reinforcement learning is not a substitute for governance but a system-level capability that requires calibrated incentives, bounded autonomy, robust evaluation, and human oversight. Sustainable adoption depends on treating adaptation as an institutional process rather than a purely algorithmic one. The conclusion emphasizes that the value of reinforcement learning in financial fraud detection lies in its capacity to learn under uncertainty, but only when embedded within accountable infrastructure, fair decision rules, and resilient operating models.

Keywords

reinforcement learning; fraud detection; financial systems; adaptive governance; machine learning infrastructure; fairness; operational resilience; policy.

1. Introduction

Financial fraud detection has become a moving-target problem. Fraud strategies evolve in response to controls, customer behavior changes across channels, payment rails expand, and economic conditions alter the distribution of legitimate and illicit activity. Conventional supervised learning models can perform well on historical data, yet they often degrade when the environment changes, when labels arrive late, or when adversaries intentionally probe decision boundaries. This has prompted interest in adaptive systems that can learn from ongoing interaction rather than depend solely on periodic retraining. Reinforcement learning offers one such perspective because it frames decision-making as sequential, feedback-driven, and oriented toward long-term outcomes rather than isolated classification accuracy [1]. In fraud detection, that framing is attractive because each decision to approve, decline, challenge, or escalate a transaction affects future information, customer trust, adversary behavior, and institutional risk.

The broader machine learning literature provides evidence that reinforcement learning can achieve high performance in complex environments when combined with function approximation and scalable computation [2], [3]. Continuous control and deep reinforcement learning methods have further demonstrated that agents can learn policies in high-dimensional settings [4]. These advances do not automatically transfer to financial fraud, because financial environments include adversarial actors, regulatory constraints, incomplete information, and severe consequences for error. However, they establish a technical basis for adaptive decision intelligence. Recent work on trust-calibrated multi-agent decision intelligence under market regime shifts [5] illustrates the importance of aligning adaptive agents with changing conditions and calibrated confidence. The key question is not whether reinforcement learning can be applied mechanically to fraud detection, but how its architecture, incentives, and governance can be designed so that adaptation improves system performance without undermining fairness, stability, or accountability.

Fraud detection research has long recognized that statistical and data mining approaches must contend with class imbalance, changing fraud patterns, and operational constraints [6], [7]. Reviews of financial fraud detection show that methods ranging from logistic regression to ensemble learning and neural networks have been applied across credit cards, insurance, telecommunications, and banking [8]. Yet many systems remain batch-oriented, with periodic retraining and manual threshold selection. Reinforcement learning introduces a different rhythm: decisions are made continuously, feedback is incorporated over time, and policy improvement becomes an operational process. This shift creates opportunities for more responsive fraud control, especially in real-time payment systems, but it also raises system-level questions about infrastructure, governance, and sustainability.

This paper argues that reinforcement learning for fraud detection should be understood as a socio-technical capability rather than a standalone algorithm. The central challenge is to design adaptive agents that can learn from changing financial environments while remaining bounded by policy, auditable by institutions, and fair to consumers. The discussion therefore moves from foundations and system context to architecture, infrastructure, governance, evaluation, and institutional learning. The aim is to provide a system-level account of how reinforcement learning can be responsibly integrated into fraud detection in changing financial environments.

2. Foundations and System Context of Adaptive Fraud Detection

Fraud detection systems operate within a complex institutional field. Banks, payment processors, merchants, card networks, and regulators all influence what counts as suspicious, what action is permitted, and what evidence must be preserved. A fraud decision is not merely a prediction about a transaction; it is an intervention in a customer relationship and a financial network. False positives can block legitimate purchases, damage trust, and generate costly manual review. False negatives can result in chargebacks, regulatory penalties, and systemic losses. The optimal operating point therefore depends on institutional risk appetite, customer expectations, and the cost structure of investigation. Reinforcement learning must be evaluated against this broader decision context rather than against classification metrics alone [6], [9].

The data environment is equally challenging. Fraud datasets are highly imbalanced, labels are often delayed or incomplete, and legitimate behavior changes because of seasonality, promotions, macroeconomic shocks, and technological adoption. Credit card fraud studies have shown that sampling strategies, feature engineering, and evaluation methods profoundly affect perceived model performance [9], [10]. Comparative studies of data mining techniques further demonstrate that no single method dominates across all fraud types and operational settings [10]. Survey work emphasizes that fraud detection requires a combination of detection accuracy, interpretability, timeliness, and adaptability [11]. Reinforcement learning can address some of these requirements by learning policies that optimize long-term expected value, but it also intensifies dependence on reward design, simulation quality, and monitoring.

Graph-based and relational approaches are especially relevant because fraud often involves coordinated behavior among accounts, devices, merchants, and geographic locations. Graph anomaly detection methods can identify suspicious structures that are invisible in isolated transaction records [15]. Fraud analytics practice similarly emphasizes social network techniques, link analysis, and entity resolution [14]. Reinforcement learning can complement these methods by learning sequential investigation policies over graph-structured evidence, such as deciding which accounts to examine next or when to escalate a case. In this sense, adaptive fraud detection is not only about classifying individual transactions; it is about managing an investigative process under uncertainty.

Market regime shifts complicate the picture further. A policy trained during a period of low fraud may behave poorly during a crisis, a new payment technology rollout, or a regulatory change. Multi-agent decision intelligence under regime shifts highlights the need for trust calibration, coordination, and context-sensitive adaptation [5]. For fraud detection, this implies that an agent should not merely maximize approval volume or minimize immediate loss; it should also maintain calibrated confidence, detect when its own assumptions have become unreliable, and defer to human review or conservative defaults when uncertainty is high. The system context therefore requires a hybrid architecture in which reinforcement learning is one component among rules, supervised models, human expertise, and governance controls.

3. Reinforcement Learning Architectures for Financial Fraud Detection

Reinforcement learning differs from supervised learning because an agent learns through interaction with an environment. It balances exploration of new actions with exploitation of known effective policies, and it optimizes cumulative reward over time [1]. In fraud detection, this can be translated into several architectural patterns. A transaction-level agent might decide whether to approve, decline, or challenge a transaction based on sequential features and estimated risk. A case-level agent might prioritize alerts for human investigators, learning

which cases are most likely to yield confirmed fraud and which can be safely closed. A network-level agent might coordinate multiple institutions or business units, adjusting thresholds and investigation capacity across a portfolio. Each pattern involves different reward structures, observation spaces, and safety constraints.

Deep reinforcement learning methods make it possible to learn policies from high-dimensional inputs such as transaction sequences, device fingerprints, merchant histories, and graph embeddings [2], [4]. However, financial fraud detection rarely offers a clean simulator in which an agent can explore freely. Real-world deployment must contend with limited exploration, because declining legitimate transactions to learn is unacceptable at scale. This has led to interest in offline reinforcement learning, batch-constrained learning, and simulation-based training. Challenges of real-world reinforcement learning include delayed rewards, non-stationarity, safety constraints, and the difficulty of reproducing deployment conditions [24]. These challenges are particularly acute in finance, where audit trails and consumer protection requirements restrict the ability to experiment.

Multi-agent reinforcement learning is relevant because fraud detection involves multiple actors with partially aligned incentives. Merchants may prefer frictionless approvals, issuers may prefer loss reduction, regulators may prioritize consumer protection, and fraudsters actively adapt to controls. A multi-agent approach can model these interactions and learn coordinated policies, but it also introduces risks of instability, collusion-like behavior, and responsibility diffusion. Trust-calibrated multi-agent decision intelligence under market regime shifts [5] suggests that agents need mechanisms for evaluating the reliability of other agents and adjusting their influence accordingly. In fraud detection, this could mean that a payment network agent learns when to trust an issuer model, when to request additional authentication, and when to escalate to a human-led process.

Reward design is a central governance issue. If the reward emphasizes immediate fraud loss reduction, the agent may over-block legitimate activity. If it emphasizes customer convenience, it may under-detect fraud. If it emphasizes investigation efficiency, it may ignore long-term reputational or regulatory costs. A well-designed reward structure should incorporate multiple objectives, including financial loss, customer friction, fairness, auditability, and systemic stability. Yet multi-objective reinforcement learning is difficult to implement and explain. Interpretability methods and inherently interpretable models become important because stakeholders need to understand why an adaptive policy chooses one action over another [16], [17]. The architecture must therefore include not only learning components but also explanation, appeal, and oversight mechanisms.

4. Data Infrastructure, Streaming, and Deployment Architectures

Adaptive fraud detection depends on infrastructure that can ingest, process, and act on data in near real time. Financial transactions generate high-volume streams that require low-latency scoring, feature computation, and decision execution. Streaming fraud detection frameworks have demonstrated that scalable architectures can process large volumes of credit card transactions and update models continuously [12]. Such systems often combine stream processing, feature stores, model serving, and feedback collection. Reinforcement learning adds further requirements because the agent must receive reward signals, update policies safely, and maintain versioned behavior for audit. The infrastructure must therefore support both fast inference and controlled learning loops.

Deployment architecture determines whether adaptation is sustainable. A centralized model may be easier to govern but slower to respond to local fraud patterns. A federated or distributed architecture may improve responsiveness but complicate consistency and accountability. Financial institutions often operate hybrid environments with on-premises systems, cloud services, and third-party vendors. This creates integration challenges involving data sharing, privacy, encryption, and contractual controls. Fraud analytics practice shows that effective systems depend on entity resolution, feature pipelines, and feedback from investigators [14]. Reinforcement learning agents must be integrated into these pipelines without bypassing data quality controls or creating unmonitored decision paths.

Monitoring is another critical infrastructure function. Fraud patterns drift, customer behavior changes, and adversarial actors probe defenses. A reinforcement learning agent may adapt in ways that are difficult to detect, especially if policy updates are frequent. Monitoring should track not only accuracy and loss but also action distributions, approval rates across demographic groups, escalation volumes, customer complaints, and model uncertainty. Safety frameworks such as AI safety gridworlds [25] illustrate the value of testing agents in controlled environments before deployment. In finance, shadow deployment, canary releases, and human-in-the-loop review can provide bounded testing grounds. These practices align with operational resilience principles that require institutions to prepare for disruptions and maintain critical services [21].

The choice of learning cadence is also consequential. Continuous learning can improve responsiveness but increases instability and oversight burden. Periodic learning is easier to audit but may lag behind fraud. A tiered approach may be preferable: low-risk decisions can be handled by fast adaptive policies, while high-risk or novel cases are escalated to conservative rules or human review. This tiered architecture reflects the reality that not all fraud decisions require the same level of autonomy. It also supports regulatory expectations for human oversight and model risk management. The infrastructure must therefore encode decision rights, escalation paths, and rollback procedures alongside reinforcement learning algorithms.

5. Governance, Fairness, Robustness, and Policy

Governance is not an external constraint on reinforcement learning; it is a constitutive part of any legitimate fraud detection system. Financial institutions are subject to model risk management, consumer protection, anti-money laundering, privacy, and operational resilience requirements. Adaptive agents that change their behavior over time can complicate compliance because the model being audited may not be the model making decisions. Governance must therefore address versioning, change control, documentation, validation, and ongoing monitoring. Principles for operational resilience emphasize the need for institutions to identify critical functions, set impact tolerances, and test recovery capabilities [21]. Reinforcement learning systems should be included in these arrangements, especially when they influence payment authorization or account restrictions.

Fairness is a central concern. Fraud detection models can disproportionately affect vulnerable groups if features correlate with protected characteristics or if enforcement patterns are uneven. Fairness and machine learning research highlights that fairness is not a single metric but a set of contested normative commitments involving allocation, representation, and due process [18]. In adaptive systems, fairness must be evaluated over time because policy updates can create feedback loops. If an agent learns to escalate certain neighborhoods or merchant categories more aggressively, it may generate more confirmed fraud there,

reinforcing its own assumptions. Governance should require disaggregated performance monitoring, appeal mechanisms, and periodic bias audits. Interpretable models and explanation tools can support these processes, though they do not replace substantive policy choices [16], [17].

Robustness and security are equally important. Fraudsters are adaptive adversaries who may deliberately manipulate features, probe decision boundaries, or imitate legitimate behavior. Reinforcement learning agents that explore can be exploited if adversaries learn their policy. Safety and robustness research emphasizes the need for conservative behavior under uncertainty, adversarial testing, and formal constraints on action spaces [25]. In financial systems, this may mean limiting the agent's authority, using ensemble or committee decisions, and requiring multi-factor authentication for high-risk actions. Robustness also includes resilience to data outages, cyberattacks, and third-party failures. The Basel Committee's operational resilience principles provide a useful frame for integrating these concerns into enterprise risk management [21].

Policy and regulation shape the space of acceptable adaptation. The European Union's proposed Artificial Intelligence Act would classify certain financial uses as high-risk and impose requirements for risk management, data governance, transparency, human oversight, and accuracy [19]. The Financial Stability Board has examined how artificial intelligence and machine learning in financial services may affect market dynamics and financial stability [20]. The National Institute of Standards and Technology's AI Risk Management Framework provides a voluntary structure for governing AI risks across design, development, deployment, and use [22]. These instruments do not prescribe reinforcement learning specifically, but they establish expectations that apply to adaptive fraud detection. Compliance must be designed into the system architecture rather than added after deployment.

6. Sustainability, Evaluation, and Institutional Learning

Sustainability in adaptive fraud detection has multiple meanings. It includes computational cost, environmental impact, organizational capacity, and long-term trust. Reinforcement learning can be computationally expensive, especially when training deep policies or running large-scale simulations. The survey literature on deep reinforcement learning highlights the resource intensity of modern methods [23]. In financial institutions, this cost must be justified by measurable improvements in fraud loss, customer experience, and operational efficiency. A system that continuously retrains but produces marginal gains may not be sustainable. Evaluation should therefore compare reinforcement learning against simpler adaptive rules, supervised models, and hybrid approaches.

Evaluation itself is difficult. Offline evaluation using historical data cannot fully capture the effects of a new policy because the data were generated by a different policy. Online evaluation through randomized controlled trials may be possible in some contexts but raises ethical and operational concerns when decisions affect customers. Simulation can help, but simulations are only as good as their assumptions about adversary behavior, customer response, and investigator capacity. Reinforcement learning research on real-world challenges emphasizes the need for benchmarks, reproducibility, and safety-aware evaluation [24]. In fraud detection, evaluation should include not only precision and recall but also approval rates, false-positive burden, time-to-resolution, complaint rates, and fairness across groups. These metrics should be monitored continuously and reported to governance bodies.

Institutional learning is a necessary complement to algorithmic learning. Human investigators, compliance officers, and customer service representatives possess contextual knowledge that may not be captured in data. Reinforcement learning agents should be designed to learn from human feedback, not to replace it. Case-based reasoning, active learning, and human-in-the-loop reinforcement learning can integrate expert judgment into policy improvement. This is particularly important for novel fraud patterns, where historical data are sparse and the cost of error is high. Research on fraud detection systems has long emphasized the importance of combining automated detection with investigator workflows [11], [13]. Adaptive agents should augment these workflows by prioritizing cases, suggesting investigative actions, and explaining their reasoning.

Cross-domain comparisons offer useful lessons. In cybersecurity, adaptive agents are used for intrusion detection, but they operate under similar constraints of adversarial adaptation and false alarms. In supply chain risk, reinforcement learning has been explored for inventory and disruption response, where delayed feedback and risk aversion matter. In healthcare, adaptive treatment policies face strict ethical and regulatory scrutiny. These domains suggest that successful deployment depends on bounded autonomy, clear accountability, and robust monitoring. For financial fraud detection, the implication is that reinforcement learning should be embedded in a broader decision intelligence stack that includes rules, supervised models, graph analytics, human review, and policy controls.

7. Conclusion

Reinforcement learning offers a powerful conceptual and technical foundation for adaptive fraud detection in changing financial environments. It can help institutions move beyond static classification toward sequential decision-making that accounts for long-term consequences, adversarial behavior, and shifting market conditions. However, the value of reinforcement learning depends less on the elegance of the algorithm than on the quality of the surrounding system. Data infrastructure, reward design, governance, fairness, robustness, and policy alignment determine whether adaptation improves fraud control or creates new risks. The analysis in this paper suggests that reinforcement learning should be deployed as a bounded, auditable, and human-supervised capability rather than as an autonomous authority.

The most promising architectures are likely to be hybrid. They combine fast adaptive policies for low-risk decisions with conservative rules and human escalation for high-risk or uncertain cases. They use simulation and shadow deployment to test policies before live use. They monitor for drift, bias, instability, and adversarial manipulation. They document changes, preserve audit trails, and provide explanations to customers and regulators. They also evaluate success using multiple criteria, including financial loss, customer friction, fairness, and operational resilience. These requirements are demanding, but they are consistent with existing expectations for model risk management and operational resilience in financial services [21], [22].

Future research should focus on safe offline reinforcement learning, multi-agent coordination under regime shifts, interpretable policy representations, and fairness-aware reward design. It should also examine the organizational conditions that enable institutions to learn from adaptive systems without losing accountability. Policy frameworks will need to evolve as reinforcement learning becomes more embedded in financial decision-making. The central lesson is that adaptation is not only a technical property of an agent; it is a property of the entire socio-technical system. Financial fraud detection will benefit from reinforcement

learning only when the system as a whole is designed to learn responsibly, govern transparently, and protect the interests of consumers and institutions alike.

References

1. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.

2. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533. https://doi.org/10.1038/nature14236

3. Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484–489. https://doi.org/10.1038/nature16961

4. Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2016). Continuous control with deep reinforcement learning. In 4th International Conference on Learning Representations. https://arxiv.org/abs/1509.02971

5. Liu, T. (2026). Trust-Calibrated Multi-Agent Decision Intelligence for Financial Risk Management under Market Regime Shifts. Available at SSRN 7307503.

6. Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. Statistical Science, 17(3), 235–255. https://doi.org/10.1214/ss/1042727940

7. Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. arXiv. https://arxiv.org/abs/1009.6119

8. 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

9. 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

10. 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

11. Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. Journal of Network and Computer Applications, 68, 90–113. https://doi.org/10.1016/j.jnca.2016.03.012

12. Carcillo, F., Dal Pozzolo, A., Le Borgne, Y.-A., Caelen, O., Mazzer, Y., & Bontempi, G. (2018). SCARFF: A scalable framework for streaming credit card fraud detection with Spark. Information Fusion, 41, 182–194. https://doi.org/10.1016/j.inffus.2017.09.005

13. Zhang, Z., Zhou, X., Zhang, X., Wang, L., & Wang, P. (2018). A model based on convolutional neural network for online transaction fraud detection. Security and Communication Networks, 2018, 5680264. https://doi.org/10.1155/2018/5680264

14. Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). Fraud analytics using descriptive, predictive, and social network techniques. Wiley.

15. Akoglu, L., Tong, H., & Koutra, D. (2015). Graph based anomaly detection and description: A survey. Data Mining and Knowledge Discovery, 29(3), 626–688. https://doi.org/10.1007/s10618-014-0365-y

16. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv. https://arxiv.org/abs/1702.08608

17. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

18. Barocas, S., Hardt, M., & Narayanan, A. (2023). Fairness and machine learning: Limitations and opportunities. MIT Press.

19. European Commission. (2021). Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). COM/2021/206 final.

20. Financial Stability Board. (2017). Artificial intelligence and machine learning in financial services: Market developments and financial stability implications. FSB.

21. Basel Committee on Banking Supervision. (2021). Principles for operational resilience. Bank for International Settlements.

22. National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0). NIST AI 100-1. https://doi.org/10.6028/NIST.AI.100-1

23. Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). A brief survey of deep reinforcement learning. IEEE Signal Processing Magazine, 34(6), 26–38. https://doi.org/10.1109/MSP.2017.2743240

24. Dulac-Arnold, G., Levine, N., Mankowitz, D. J., Li, J., Paduraru, C., Gowal, S., & Hester, T. (2021). Challenges of real-world reinforcement learning: Definitions, benchmarks and analysis. Machine Learning, 110(9), 2419–2468. https://doi.org/10.1007/s10994-021-05961-4

25. Leike, J., Martic, M., Krakovna, V., Ortega, P. A., Everitt, T., Lefrancq, A., Orseau, L., & Legg, S. (2017). AI safety gridworlds. arXiv. https://arxiv.org/abs/1711.09883