FEDERAL RESERVE BANK
PHILADELPHIA

APRIL 2026

AI-Enabled Fraud Is On the Rise — Here's How to Beat It

Julapa Jagtiani
Federal Reserve Bank of Philadelphia

Raghavendra Rau
University of Cambridge

Shivam Srivastava
Indian School of Business, Hyderabad

Siddhartha Patala
Indian Institute of Management (IIM), Visakhapatnam

AI-Enabled Fraud Is On the Rise — Here's How to Beat It

Julapa Jagtiani*
Federal Reserve Bank of Philadelphia

Raghavendra Rau
University of Cambridge

Shivam Srivastava
Indian School of Business, Hyderabad

Siddhartha Patala
Indian Institute of Management (IIM), Visakhapatnam

March 26, 2026

Executive Summary

A February 2024 deepfake scam in Hong Kong, where an employee wired $3.2 million to fraudsters posing as executives, highlights a critical reality: AI-enabled fraud is on the rise. In the past year, about 79 percent of financial institutions experienced fraud attempts. As AI-driven attacks outpace traditional defenses, the core challenge is no longer flawless prediction, but the ability to buy time to respond before damage occurs. We argue that AI's primary value in fraud prevention is not merely speed but its capacity to detect uncertainty early and deliberately slow down risky transactions, expanding the window for human judgment. This challenges the conventional "faster is better" mantra. We propose a structured, 90-day implementation plan to fight against growing AI-enabled fraud. By deliberately introducing friction and escalating uncertainty earlier, our approach could prevent more losses than perfect, real-time prediction alone.

Keywords: AI, cybersecurity, bank risk, cyber risk management, fintech
JEL Classification: G21, G32, M15

---

*Please direct correspondence to Julapa Jagtiani, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106; Email: Julapa.Jagtiani@phil.frb.org; Phone: 267-275-6253.

The views expressed in this article are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

In February 2024, an employee at a multinational engineering firm in Hong Kong wired nearly 25 million Hong Kong dollars (HKD), equivalent to $3.2 million, after joining what looked like a routine video call with senior executives.[1] The faces and voices matched his colleagues — but they were fake. By the time anyone noticed, the money had moved beyond recovery.

This was not an isolated incident. AI-enabled fraud has been on the rise.[2] In October 2021, fraudsters in the United Arab Emirates used an AI-generated voice to impersonate a senior executive, convincing a bank manager to approve a transfer of $35 million.[3] The manager had spoken to the real executive before, but the synthetic voice was so convincing, it raised no red flags. Cases like these keep rising as cyberattacks now unfold at machine speed (i.e., in seconds or minutes), while many controls still assume attacks will take hours or days to complete the transaction. Insights from IBM's analysis of deepfake-driven cybercrime show that AI-generated voices, videos, images, and text are increasingly used in real-time interactions, such as video calls, contacts with call centers, and internal help desk chats, in which decisions are made quickly and with limited verification windows.[4]

In traditional wire fraud schemes, attackers relied on prolonged email exchanges, manual approvals, and the slow, incremental building of trust with the target, with fraudulent fund transfers typically occurring over two to three days. In contrast, deepfake-enabled fraud has drastically accelerated, with attack cycles often completing in under 90 minutes and some executive impersonation cases triggering fraudulent fund transfers within 10 to 30 minutes of initial contact.[5]

---

[1] See Heather Chen and Kathleen Magramo, "Finance Worker Pays Out $25 Million after Video Call with Deepfake 'Chief Financial Officer'," CNN, February 4, 2024, https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk.

[2] See Derek Manky and Gil Baram, "Beyond Phishing: Exploring the Rise of AI-Enabled Cybercrime," University of California Berkeley Center for Long-Term Cybersecurity, January 2025, cltc.berkeley.edu/2025/01/16/beyond-phishing-exploring-the-rise-of-ai-enabled-cybercrime/ See also Amanda Gerut, "Consumers Lost $12.5 Billion to Fraud Last Year, and AI-Powered Scams Are Set to Explode in 2026, Experian Warns," *Fortune*, January 13, 2026, fortune.com/2026/01/13/ai-fraud-forecast-2026-experian-deepfakes-scams/, and PWC, "The Fraud Trend to Watch in 2026 and Beyond. The Era of Deepfakes and Synthetic Identities," February 16, 2026, www.pwc.com/cz/cs/blog/rizeni-rizik/the-fraud-trend-to-watch-in-2026-and-beyond.html.

[3] See Thomas Brewster, "Fraudsters Cloned Company Director's Voice In $35 Million Heist, Police Find," *Forbes*, October 14, 2021, www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/; see also Alvarez Technology Group, "Cybercriminals Use AI to Steal Large Sum from UAE Bank," November 4, 2023, www.alvareztg.com/uae-bank-deepfake/.

[4] See Srini Tummalapenta, "How a New Wave of Deepfake-Driven Cybercrime Targets Businesses," *Think*, IBM, accessed on February 16, 2026, www.ibm.com/think/insights/new-wave-deepfake-cybercrime.

[5] See Reshma Sunil, Parita Mer, Anjali Diwan, Rajesh Mahadeva, and Anuj Sharma, "Exploring Autonomous Methods for Deepfake Detection: A Detailed Survey on Techniques and Evaluation," *Heliyon* 11:3, February 2025, e42273, www.sciencedirect.com/science/article/pii/S240584402500653X?via%3Dihub, and Satish Lalchand, Val Srinivas, Brendan Maggiore, and Joshua Henderson, "Generative AI Is Expected to Magnify the Risk of Deepfakes and Other Fraud in Banking," Deloitte Center for Financial Services, May 29, 2024,

This rapid compression of the attack life cycle is driven by real-time AI tools that eliminate the need for lengthy, manual coordination.

The core managerial challenge in the age of AI-driven fraud is that latency has replaced model accuracy as the primary vulnerability. In an environment in which fraud executes faster than traditional governance processes can respond, the safeguard is not just the ability to identify fraud accurately but also the ability to create a bottleneck, buying time for firms to respond before significant damage occurs. This shift has redefined the core managerial challenge from improving detection accuracy to minimizing system latency.

Why Is This Happening Now?

Generative tools have slashed the cost of creating convincing deepfakes, while faster payment rails have reduced the friction that used to give teams time to verify. Legacy controls built for slower threats are ineffective when a plausible request can be created, routed, and executed in minutes. And the scale is significant.[6]

According to the 2025 Association for Financial Professionals (AFP) Payments Fraud and Control Survey, 79 percent of financial institutions worldwide were victims of attempted or actual payments fraud in 2024, with recovery outcomes becoming increasingly dire.[7] Only 22 percent of organizations were able to recover 75 percent or more of the funds lost to fraudulent payments in 2024, representing a sharp decrease from 41 percent in 2023. This reflects the widening gap between human-paced organizational responses and AI-induced cyberattacks occurring at machine speed.

An IBM report highlights that the barrier to entry for deepfakes has collapsed. This technology has been democratized — transitioning from a niche, high-effort capability into an accessible tool for nontechnical attackers.[8] Freely available data sets, lightweight neural networks, and cloud computing now enable high quality deepfake audio and video to be generated in nearly real time. Europol reports the rise of "deepfake-as-a-service," with threat actors willing to pay $16,000 for custom deepfake services — significantly reducing technical barriers while maximizing

---

www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html.

[6] See Carter Pape, "$300 AI Tool Kits Let Criminals Bypass Bank Security," *American Banker*, March 25, 2026, www.americanbanker.com/news/300-ai-toolkits-let-criminals-bypass-bank-security.

[7] See Association for Financial Professionals, "2025 AFP Payments Fraud and Control Survey Report," accessed on February 16, 2026, www.financialprofessionals.org/training-resources/resources/survey-research-economic-data/details/payments-fraud.

[8] See Tummalapenta "How a New Wave." Specifically, the article states that "The barrier to entry for bad actors is lower now than before. Tools allowing the creation of deepfakes are cheaper and more accessible than ever, giving even the users with no technical know-how the chance to engineer sophisticated, AI-fueled fraud campaigns."

the potential illegal financial upside. We believe that while firms cannot eliminate fraud entirely, they could detect suspicious activities quickly enough to slow execution and intervene before transactions become irreversible.

What Works in Practice?

Several banks now use real-time authorization tools that score transactions before funds move. Mastercard's Consumer Fraud Risk system runs its AI models in the authorization flow to evaluate transaction context and generate live risk signals for sending banks before the execution of payment transactions.[9] The sending bank could then pause or block the transaction before execution.

Related to this, based on 2024 data from UK Finance and associated industry reporting, the effectiveness of authorized push payment (APP) fraud prevention has shifted from relying solely on model accuracy to creating actionable time for human intervention. In the United Kingdom, APP fraud losses fell 2 percent to just over £450 million in 2024, while the number of APP fraud cases dropped by 20 percent, representing the lowest level since 2021.[10] These reductions followed sustained industry investment in real-time detection, behavioral monitoring, and intervention mechanisms designed to interrupt transactions before execution. In fast-moving attacks, that delay matters more than a marginal improvement in prediction.

Triage at Scale with a Human in the Loop

High-volume alert queues overwhelm investigators and generate fatigue. An AI assistant turns raw data into actionable insights — it summarizes cases, clusters similar alerts, identifies trends, generates detailed reports, and proposes a next action that lets senior analysts focus on the riskiest items for timely decisions. The success metric is straightforward: fewer low-value escalations and shorter median case times without an increase in missed fraud.[11]

Consider a hypothetical regional bank processing 800 alerts daily across cards, wires, and digital channels, with a baseline average handling time of 15 minutes per alert and roughly 90 percent of those alerts ultimately closed as false positives. After deploying an AI assistant that

---

[9] See Mastercard (2024) "Mastercard Expands First-of-Its-Kind AI Technology to Help Banks Protect More Consumers from Scams in Real Time," press release, September 24, 2024, www.mastercard.com/us/en/news-and-trends/press/2024/september/mastercard-expands-first-of-its-kind-ai-technology-to-help-banks-protect-more-consumers-from-scams-in-real-time.html.

[10] See UK Finance, "Fraud Continues to Pose a Major Threat with Over £1 Billion Stolen in 2024," press release, accessed on February 16, 2026, www.ukfinance.org.uk/news-and-insight/press-release/fraud-report-2025-press-release.

[11] See, for example, LeewayHertz, "AI in Fraud Detection: Use Cases, Architecture, Benefits, Solution and Implementation," web page, accessed on March 20, 2026, www.leewayhertz.com/ai-in-fraud-detection/.

auto-clusters related alerts into cases, generates concise summaries with key transactions and customer context, and recommends whether to auto-close, request more information, or escalate, the bank could reduce the number of alerts requiring manual review by 50 percent[12] from 800 to less than 400 alerts per day, and can cut the average handling time from 15 minutes to 9 minutes, while still capturing essentially all previously detected fraud.[13]

Manage Attention Shocks

Fraud surges often coincide with narrative spikes (e.g., during highly emotional events or other urgent situations) that distract staff and customers.[14] Market events often trigger a surge in social media–driven scams and other fraudulent activities, driving up spoofed payment requests and mule account (a bank account that attackers use to receive, move, or withdraw stolen funds) recruitment.[15]

During the GameStop episode in early 2021, when retail trader Keith Gill ("Roaring Kitty") catalyzed a short squeeze that drove extreme market volatility.[16] The surge in attention was accompanied by a documented rise in fraud and financial crime attempts. Cybersecurity firms and financial regulators reported a sharp increase in equity-themed phishing, impersonation, and account-takeover attempts during the January 2021 meme-stock period. Industry threat reports indicate that stock market–related phishing activity increased by 150–200 percent over baseline levels, with GameStop, AMC, and Robinhood consistently cited as primary social-engineering lures. Attackers exploited urgency, trading halts, and price volatility to induce credential theft and fraudulent fund transfers, particularly targeting retail investors.

A similar pattern reemerged in May 2024, when Gill returned to social media after nearly three years of inactivity.[17] In 24–48 hours, social-listening and threat-intelligence platforms

---

[12] This 50 percent statistic has been observed in practice: see, for example, DataBahn, "Reduced Alert Fatigue: 50% Log Volume Reduction with AI-Powered Log Prioritization," April 7, 2025, www.databahn.ai/blog/log-prioritization-volume-reduction-microsoft-sentinel.

[13] See also a case study by WorkFusion, "US-Based Bank Streamlines Real-Time Payments Compliance and Reduces Monetary Waste in Level 1 Alerts Review," case study, , 2024, www.workfusion.com/wp-content/uploads/2024/01/US-based-bank-streamlines-real-time-payments-compliance-WorkFusion.pdf.

[14] The term "narrative spike" refers to a sudden, intense surge of interest, heightened attention, or emotional arousal. During the COVID-19 pandemic, attackers launched phishing and malware attacks to take advantage of fear during the unprecedented business lockdown by sending emails (e.g., disguised as updates from the World Health Organization, Centers for Disease Control, or Internal Revenue Service) containing ransomware attachments to steal credentials from remote and furloughed workers.

[15] Mule accounts are often associated with criminal activities, and they may also be overseas (offshore).

[16] See Greg Iacurci, "GameStop Mania Fed Off Investor Angst. Experts Say that Unease Still Fuels 'Gamblifying' of Investing," CNBC, January 29, 2026, www.cnbc.com/2026/01/29/gamestop-meme-stocks-retail-investors-wall-street-young-investors.html.

[17] See Krystal Hur, "GameStop Meme Lord Behind Stock's Wild Moves Reveals Himself After Three Years," CNN, June 7, 2024, www.cnn.com/2024/06/07/investing/roaring-kitty-keith-gill-gamestop.

recorded a sudden, order-of-magnitude increase in GameStop-related online discourse, followed by a corresponding rise in scam activity.[18] Cybersecurity vendors documented more than a 300 percent increase in meme-stock-themed phishing messages and fraudulent domains, many impersonating brokerages, trading platforms, and regulators. These episodes show that narrative shocks function as predictable accelerants of financial crime, compressing fraud timelines and increasing attack success rates by exploiting heightened attention, emotional arousal, and information asymmetry.

Some institutions are testing narrative monitors that flag these intervals and trigger tighter verification rules for high-risk requests. The goal is not to predict the market but to increase operational readiness — knowing when to temporarily elevate scrutiny because fraudsters are likely exploiting the same events.

Avoid Confident-but-Wrong Automation

When models trained on yesterday's patterns are faced with a regime shift, errors correlate. Moreover, using multiple AI systems can amplify shared assumptions and produce large losses. Zillow's iBuying episode offers a cautionary example from outside finance.[19] In 2021, Zillow Offers relied on machine learning models trained primarily on historical transaction data from a prolonged period of rapid home price appreciation. These models assumed short-horizon price predictability, continued upward momentum, and stable market liquidity. They were designed to forecast resale prices three to six months into the future and to scale purchasing volume aggressively based on those forecasts. Critically, Zillow deployed multiple interdependent models for valuation, bidding, inventory accumulation, and resale timing that were trained on the same data distributions and embedded the same market assumptions. As a result, the system exhibited correlated errors rather than independent checks.

When housing markets began to slow and regional volatility increased in mid-2021, the models systematically overestimated future prices. Zillow purchased homes near local price peaks, and operational constraints in renovation and closing delayed resale. Because the models failed in the same direction, errors compounded across the pipeline. Zillow accumulated a large inventory of homes that could only be sold at discounts, resulting in quarterly net loss of $328 million. In November 2021, Zillow exited the iBuying business, laid off about 25 percent of its workforce, and

[18] See Ben Bain and Daniel Avis, "SEC Hunts for Fraud in Social-Media Posts Hyping GameStop," Bloomberg, February 3, 2021, finance.yahoo.com/news/sec-hunts-fraud-social-media-192424991.html.
[19] See Padma Susarla, Dexter Purnell, and Ken Scott, "Zillow's Artificial Intelligence Failure and Its Impact on Perceived Trust in Information Systems," *Journal of Information Technology Teaching Cases*, 2024, doi.org/10.1177/20438869241279865.

began liquidating its remaining inventory at expected losses of 5 percent to 7 percent per home. The company ultimately reported $881 million in losses for 2021.

Postmortem analyses emphasized that the failure did not stem from insufficient data or weak modeling techniques but instead stemmed from a system-level design flaw in which multiple models shared the same assumptions and therefore failed simultaneously when market conditions shifted. In fraud prevention, the parallel risk is direct. If transaction scoring, voice authentication, and anomaly detection systems are all trained on the same historical data and encode similar assumptions, they will miss the same emerging threat patterns.

Redundant AI models without epistemic diversity increase fragility, rather than robustness. To achieve true robustness, the solution is not to deploy more models but to design systems to embrace model disagreement as a feature of uncertainty **—** models are expected to disagree under uncertainty. Instead of treating divergent model outputs as errors to be minimized, firms should use disagreement as a control mechanism to trigger friction and necessary human intervention.

What Leaders Could Accomplish in 90 Days

We propose a structured, 90-day implementation plan based on recent industry best practices for effective control to fight against growing AI-enabled frauds.

Days 1–25: Adopt (Focus on Assisted Fraud-Alert Triage)

A successful AI adoption begins with clear accountability and a specific goal: migrating from manual processes to AI-driven triage to reduce operational workloads.

Consider a regional bank's pilot program overseen by the head of fraud operations, aiming to reduce manual investigation of false-positive transactions, i.e., legitimate transactions that are incorrectly flagged as fraud. By utilizing historical alerts from the previous quarter, the bank could introduce an AI tool that summarizes cases and provides a binary recommendation — escalate or close — to aid human decision-making. Consequently, for a workload of 800 daily alerts, average review times per case would shrink from 15 minutes to 9 minutes because key transaction details are summarized up front. Additionally, the initiative is projected to reduce the escalation rate by approximately 25 percent, resulting in streamlined investigation queues and enhanced efficiency.

If the AI tool fails to exceed pre-pilot baseline or reduce false positives within four weeks, the AI-assisted process would automatically switch to read-only mode (and return to its manual systems until the AI-assisted tool has been recalibrated). Performance would be tracked weekly via three CFO-reported metrics: fraud escalation rate, average time to decision (or median time it takes in each case from alert to decision), and false negatives (number of confirmed fraud cases that were initially incorrectly identified as legitimate and recommended for closing).

In this hypothetical case, early issues found were related to international wires, causing the team to temporarily limit the scope of the AI-assisted tool to only domestic transactions while it worked to retrain/recalibrate the AI process. With the recalibrated AI systems, the bank reported a lower median case time, a lower number of false positives (low-value escalations were reduced by 25–30 percent), and a lower number of false negatives (true fraud detection was improved or remained unchanged).

Days 26–65: Adapt (Move to Real-Time Sensing)

At this stage, the bank would move forward from the previous *reactive* mode of using fraud alert triage model to a *proactive* mode of using real-time sensing model and dynamic dashboards to visualize fraud trends as they happen. The goal at this stage is to leverage AI to detect and stop fraud as it occurs, rather than investigating it after the fact. This shift requires moving from batch processing of alerts to streaming data analytics and automated, AI-driven decision making — with the ability to instantly block suspicious transactions upon detecting high-risk signals.

Under the chief risk officer's responsibility, anomaly signals should be embedded into existing treasury and fraud dashboards, not into a separate tool. Key signals like transaction velocity (daily payment volume), counterparty risk (new payee, high risk jurisdiction, suspicious category), and payment size relative to historical patterns should be added to the current treasury dashboard. These can be distilled into a traffic-light framework: green for routine activity, yellow for moderate risks such as slight spikes in volume or new payees, and red for high-risk anomalies such as extreme amounts or high-risk jurisdictions. This keeps the control simple, visible, and tied directly to the payment release workflow.

At this stage, two metrics should be tracked weekly. First, median anomaly time to decision, which is the time from when a payment is flagged yellow or red to when the treasury or fraud team takes a clear action (approve, reject, escalate). The goal should be to reduce the time to decision from hours to under 15 minutes by implementing a prominent red flag system and defining clear, automated escalation paths. Second, the share of events with complete logs, which is the percentage of high-risk flagged payments that have a documented reason for the anomaly, a record of the callback (who was called, number used, outcome), and the final decision. The goal should be to increase the share of complete logs from 60–70 percent to at least 95 percent within 60 days by reengineering the payment workflow to make callbacks and loggings a synchronous, mandatory step in the payment workflow, rather than an optional process.

To proactively test the resilience of AI-assisted control systems against increasingly sophisticated deepfake and other AI-enabled payment fraud, AI-powered social engineering drills may be conducted. This would empirically validate our authentication protocols against synthetic

voice and video attacks. The goal of this exercise is to test whether staff follow callback rules (using a pre-verified number on file, not the number from the suspicious call or email), stick to authorization thresholds even when a senior executive appears to request payment, and enforce step-up authentication for urgent, high-value payments.

All gaps exposed by this exercise, such as weak callback rules, overreliance on seniority, and the absence of step-up controls in an emergency, must be remedied and retested within 30 days. A concrete example that works in practice is a single stoplight indicator in the treasury dashboard that turns yellow or red when a payment's velocity, counterparty risk, or size relative to history is elevated. When the yellow or red sign appears, treasury staff must call the requester using a verified number on file, not the number in the email or call, and log the outcome before releasing the payment. This adds about three minutes to suspicious transactions but could catch deepfake-enabled impersonation attempts almost immediately, preventing multimillion-dollar losses.

Days 66–90: Institutionalize (Build Governance That Lasts)

Institutionalizing AI governance is the final, essential step in moving towards a sustainable, trustworthy, and proactive anti-fraud framework. The goal in this step is to ensure that the bank's AI-enabled tools (implemented between Day 1 and Day 65) would be used responsibly, ethically, and continue to adapt in response to evolving fraud tactics. To formalize AI-enabled fraud defenses, the bank would follow these best practice step-by-step governance structure.

First, establish a dedicated *AI governance council*, chaired by the CFO, with core members including the chief risk officer, chief data officer, and chief information officer.[20] The council would meet monthly — approximately 90 minutes. The council provides executive oversight for high-impact AI, sets boundaries, and ensures risk management and compliance are embedded in the design and operation.

Second, the bank would designate a *primary model owner* for each high-impact model (such as fraud, payments, and onboarding) to drive accountability. At every monthly meeting, the model owner must bring to the table a risk-focused update, moving beyond simple status reports. Monthly owner mandated reports would include a review of the previous month's error logs and any drift in key metrics (such as rising false positives or falling fraud capture), and one specific question they cannot answer (such as "What edge case keeps us awake?" or "Where would this fail silently?").

Third, the bank would set up a *quarterly challenge session* in which compliance and finance jointly review errors and drift. Attendees might include council members, model owners, the head

[20] In general, the chief information officer (CIO) manages technology infrastructure and IT operations, while the chief data officer (CDO) focuses on data strategy, governance, and turning data into business value.

of fraud, and the chief compliance officer. They would review the error log and performance trends for each model and ask sharp questions. When was the last false negative that led to a near miss or loss? How have false positives changed? Has performance weakened in any segment, such as new customers or high-risk geographies? Is the override rule being followed in practice?

Fourth, if a model is underperforming, the AI governance council would have *clear options*, such as to impose tighter controls (e.g., keep it advisory-only, add a mandatory human check, cap risk scores), to pause new use cases, or to require model retraining and revalidation by a fixed date. The chief compliance officer would have veto power on any automation that touches customer funds.

Fifth, the bank would publish a *one-page playbook* that clearly states the human override boundary and what specific activity should never be automated. Examples of these boundaries could include that any payment over $500,000 requires a voice call using a number on file, not the number on the request; that any change to beneficiary details for such a payment must have a second approval and a call to the verified number; or that certain actions must be explicitly excluded from automation, such as approvals for new high-risk jurisdictions, permanent high-value limits, and overriding fraud blocks without a documented reason and approval.

Sixth, it is important for the bank to define clear, bounded *audit scopes*, such as "all payments over $100,000 reviewed within 48 hours," instead of generic, high-level, or unmeasurable promises. The goal is to produce defensible, actionable evidence for the board and regulators. This operational shift introduces deliberate friction. The trade-off is intentional. Modest upfront friction enables earlier detection, limits downstream losses, strengthens regulatory posture, and protects customer trust on a scale.

Once again, we stress that when AI becomes the operating layer for fraud prevention, speed is no longer the only priority.[21] Context and behavioral validation take precedence. Treasury and payment workflows may slow down by design as transactions are evaluated against customer baselines, cross-channel activity, and real-time risk signals. Legitimate high-value payments that once cleared in a few minutes may now take 15 to 20 minutes for verification. Some customer dissatisfaction may be unavoidable.[22]

It is also important to note that AI-driven fraud detection may not necessarily reduce workload in the first year. Institutions that deploy AI as a real-time fraud operating layer often

---

[21] See Abhinn Goswami (2026) "If Fraud Risk Had a Speed Limit, It's Been Removed. AI Is Not Just a Threat — It's a Force Multiplier," *Insights from the Frontlines*, January 14, 2026, www.linkedin.com/pulse/fraud-risk-had-speed-limit-its-been-removed-ai-just-threat-goswami-e4igc/

[22] See Technology Mindz, "The Silent Threat in Banking Operations: Why Fraud Detection Using AI Is Non-Negotiable," *Technology Mindz Blog*, accessed on February 16, 2026, technologymindz.com/the-silent-threat-in-banking-operations-why-fraud-detection-using-ai-is-non-negotiable/

report a temporary increase in fraud operations costs of approximately 10–15 percent in the first six to 12 months. The increased costs are related to initial rise in false positive, dual-system redundancy during initial adoption, integration and training costs, etc. The cost curve would bend downward after the first year as the model matures, false positives decline, automation absorbs routine decisions, and teams shift from reactive response to early risk prevention.[23]

Research shows that financial institutions incur more than four times the value of the fraudulent transaction in total cost. LexisNexis (2025) reports that the total cost of fraud has been rising in recent years — for financial institutions in North America, every $1 lost to fraud resulted in an average total cost of $4 (as of 2021) and $5.75 (as of 2025). [24] These costs include internal labor, legal fees, recovery efforts, and reputational damage (owing to loss of customer trust and inability to attract new business or talent).[25] Preventing fraud is therefore not only about saving simply the transaction value but also about avoiding a cascading cost structure that rapidly compounds beyond the initial loss.

The Managerial Mindset That Stays

The goal is not to eliminate human judgment but to earn time for human judgment on the small subset of cases that matter. This requires a shift in how success is measured. Rather than asking if you caught every fraud, you ask whether you can prove that your systems work. The evidence is in your decision logs, your challenge sessions, and your ability to show that when the model disagrees with itself, humans step in. Under this approach, AI becomes less a bet on flawless prediction and more a failsafe discipline to contain losses when attack speeds outpace human intervention.

[23] There is also the organizational cost of maintaining discipline. Fraud teams will push back on delays. Business units will want exceptions. Executives will ask why legitimate transactions are getting flagged. The governance structure needs to be strong enough to hold the line when convenient workarounds emerge.

[24] LexisNexis Risk Solutions, LexisNexis True Cost of Fraud[T] Study 2025 North America – Financial Services & Lending," Alpharetta, GA: LexisNexis Risk Solutions, 2025, risk.lexisnexis.com/-/media/files/financial%20services/research/lnrs_true_cost_of_fraud_study_2025_north_america_financial_services_lending_v2.pdf.

[25] See William C. Johnson, Wenjuan Xie, and Sangho Yi (2014) "Corporate Fraud and the Value of Reputations in the Product Market," *Journal of Corporate Finance* 25, April 2014, pp. 16–39. See also First Financial Bank (2026) "Reputational Damage: When Fraud Causes More than Financial Loss," *Flourish with First Blog*, www.bankatfirst.com/business/resources/flourish/fraud-and-reputational-damage.html, and Discover Global Network (2025) "The Role of Fraud Risk Prevention in Reputation Management," *Insights*, November 26, 2025, insights.discoverglobalnetwork.com/insights/fraud-risk-prevention-in-reputation-management.