Financial Services Sector Coordinating Council
for Critical Infrastructure Protection and Homeland Security

AI-Generated Fraud in the Financial Sector:
Threat Categories and Defense Strategies

This document is the product of the Enhanced Artificial Intelligence Fraud Workstream, established by the Financial Services Sector Coordinating Council (FSSCC) as part of the Financial and Banking Information Infrastructure Committee (FBIIC)-FSSCC Artificial Intelligence Executive Operating Group (AIEOG). The workstream, led by the Financial Services Information Sharing and Analysis Center (FS-ISAC), the American Bankers Association (ABA), and the Bank Policy Institute (BPI), includes representatives from the FBIIC and FSSCC, as well as financial institutions of all sizes and criticality.

Financial Services Sector Coordinating Council
for Critical Infrastructure Protection and Homeland Security

AI-Generated Fraud in the Financial Sector

Threat Categories and Defense Strategies

Contents

Executive Summary 2

Overview 3

Two Categories of AI-Related Attacks 4

AI Fraud Attack Landscape 5

Tactics & Deployments 5

Impostor Targets 6

Education & Awareness 7

The Goal of Fraud Education 8

Internal Education Strategy 8

External Education Strategy 10

Campaign Design Principles 11

Example Delivery Channels & Communication Tactics 12

Incident Response & Operational Reporting 13

Deepfake Response 15

Controls & Technology Response 16

Representational Outcomes for Consideration 16

Collaborative Efforts to Prevent & Disrupt Fraud 20

The Need for Tracking AI-Enabled Scam Claims 22

Why It Matters 22

Challenges in Capturing This Data 22

Recommendations 23

AI Companies Helping to Fight Fraud 23

Conclusion 24

References 25

Executive Summary

Artificial Intelligence (AI) has not created new categories of fraud – but it is fundamentally reshaping the speed, scale, and credibility of existing scams, allowing criminals to produce flawless phishing messages, convincing deepfake voices and videos, synthetic identities, and adaptive social-engineering scripts at industrial scale and at minimal cost. Regulators have already taken note, flagging deepfakes as an emerging fraud vector for Suspicious Activity Reporting (SAR).

Impersonation has emerged as the dominant attack pattern. Threat actors are successfully counterfeiting executives, financial institution staff, trusted vendors, advisors, and even the family members of financial institution (FI) employees and customers to authorize payments, redirect funds, and extract credentials – sometimes resulting in single-event losses measured in millions of dollars.

The paper distinguishes two AI-driven threat categories:

- AI used to deceive humans through phishing, smishing, deepfake audio and video, synthetic personas, and adaptive scripts that exploit trust.
- AI used to defeat identity proofing and authentication controls, including deepfake attacks against Know Your Customer (KYC) controls, biometrics, and onboarding systems.

The work provides pragmatic recommendations on:

- Recalibrating, stress-testing, and incrementally enhancing existing controls to account for AI's impact on realism, speed, and scale.
- Modernizing internal and external education strategies, such as embedding timely in-app warnings, inserting intelligent friction, and evolving incident response to handle consumer-initiated fraud at volume.

Success will depend on tighter cross-functional coordination, better measurement of AI involvement in scams, expanded intelligence sharing within the financial sector and with others, and deeper collaboration with AI technology providers. The challenge is serious but manageable. With disciplined execution, incremental enhancements, and ecosystem-level cooperation, AI-enabled fraud is a fight the financial services sector can win.

Overview

The criminal use of AI, particularly generative AI (GenAI), to enhance fraud and scam tactics is not new. However, the speed, sophistication, and scale at which threat actors can execute attacks are rapidly accelerating. Off-the-shelf AI tools can create flawless phishing lures, deepfake voices and videos, and bespoke scripts that adapt in real time to a victim's responses. That makes it harder for consumers to spot red flags and gives attackers more "shots on goal" per hour than ever before. US regulators have likewise flagged deepfakes as an emerging fraud vector for SAR.[1]

Impersonation attacks are increasingly common. Many executives, financial institution staffers, vendors, and their family members have been the targets – or impersonated subjects – of convincing voice and video deepfakes used to authorize payments, redirect funds, or extract credentials. Some of these frauds have resulted in single-event losses of millions of dollars.[2] These incidents illustrate how AI amplifies long-standing social-engineering tactics (urgency, authority, fear) and compresses the time required to execute multi-party deception.

> Deloitte projects that US fraud losses tied to GenAI activity could rise from ~ $12.3B US in 2023 to nearly $40B by 2027, driven by advancements in automation and increased credibility.
> Deloitte Center for Financial Services

FIs should plan and adapt now. As these attacks evolve, a primary concern is that Threat Actors (TAs) will direct some of this AI-enhanced fraud at consumers, who will believe they are acting responsibly in response to what appears to be a legitimate FI request, helping a family member, or following security guidance when, in reality, TAs are socially engineering the victim by criminal use of augmented AI tools. Many of the protective controls in these scenarios reside on the consumer's devices or in the consumer's decision-making moment, limiting the institution's direct ability to intervene.

This dynamic calls for a different institutional mindset that empowers consumers through continuous education, designs operations and communications to anticipate deception, modernizes incident response to include consumer-initiated fraud events, and evolves technological strategies to detect and disrupt scams earlier in the engagement cycle.

Consumers will face greater difficulty distinguishing real from fake as AI improves the *quality* of deception (fewer grammatical errors, highly tailored pretexts, realistic synthetic media). Federal Trade Commission (FTC) data already show rising consumer losses in "imposter scam" categories, consistent with more convincing outreach at scale.[3] To counter these threats, it is clear that public education must be refreshed accordingly, with more straightforward rules of thumb and in-app warnings delivered at the moment of risk.

Two Categories of AI-Related Attacks

For clarity, we distinguish two categories of AI-generated attacks:

1. **AI to deceive consumers and employees** (the focus of this paper) that primarily exploits human trust and decision-making. Empirical research indicates that AI enhances both the quantity and quality of phishing, while real-world cases demonstrate the scale of deepfake-enabled payment fraud.[4] These attacks include:

   - GenAI-written phishing and smishing
   - Deepfake voice/video impersonations of high-profile or business leaders, FIs, vendors, or family members
   - Adaptive social-engineering scripts

2. **AI to defeat identity proofing and authentication controls**

   - Deepfake or selfie injection against onboarding KYC processes
   - AI-manipulated documents
   - Voice/face clones attacking call-center or device biometrics
   - Agentic AI used for credential-stuffing or automated account takeover

     ➢ The Financial Crimes Enforcement Network (FinCEN) has already issued an alert instructing FIs to tag deepfake-related fraud in SARs, underscoring the threat profile of control evasion.[5]

Guidance concerning the use of AI to defeat identity proofing and authentication controls will be addressed in a companion paper.

Wholly new fraud types are not likely in the short term, but AI will materially increase the velocity and believability of existing fraud types. That reality argues for pragmatic, near-term changes in training, processes, and controls – paired with better measurement of AI involvement – to manage loss curves. At the same time, longer-cycle technology and policy responses must mature.

AI Fraud Attack Landscape

To understand the increase in AI-enhanced fraud, it is necessary to examine TAs' tactics and primary targets.

Tactics & Deployments

> GenAI can easily generate the well-known toll payment phishing scam and other fraud campaigns. RCS and iMessage techniques are also circumventing telecommunications industry controls.

- **Generating text and image content**: TAs can use GenAI to create tailored emails, instant messages, and images as bait to hook potential scam victims, for example, in phishing and smishing attempts or fraudulent advertisements.
- **AI-enabled chatbots**: Chatbots have the potential to scale fraudsters' ability to contact victims. Little human involvement is necessary – a single chatbot can deliver the same scam volume that once required a call center of people.
- **Deepfake video**: TAs are already using deepfake videos as lures. Some deepfakes serve as 'clickbait' linked to malicious websites that harvest card payment details for subsequent use in payment fraud.
- **Voice cloning – scams and voice ID**: Deepfake technology requires less than two hours of audio content to clone a voice, and the degree of accuracy is increasing while the requirement for training data is decreasing. In many use cases, only a few seconds of training audio are sufficient for a successful attack.
- **AI-generated documentation**: TAs are creating documents (such as loan documentation) from existing documents.
- **Synthetic personas and fake profiles:** GenAI can create realistic digital identities – complete with headshots, bios, and social-media history – to build credibility, infiltrate organizations, or defraud consumers.
- **AI-assisted reconnaissance and target profiling:** Tools can scrape and summarize victims' digital footprints (emails, social posts, corporate bios) to craft hyper-personalized scam content.

- **Scam amplification via AI content farms:** Generative tools create massive volumes of scam articles, fake "consumer reviews," or search engine-optimized content that makes fraudulent websites rank higher and appear credible.
- **AI-driven misinformation and investment scams:** Fake "news videos" or AI-generated influencer clips push crypto or investment fraud schemes, lending them the appearance of legitimacy.

Many of these attacks invoke a common theme: they create a sense of urgency that drives the victim to respond quickly and bypass traditional controls. When individuals perceive time is short – as they do under a tight deadline or in an emergency – they may feel pressured to make quick decisions. This pressure can cause them to rely on snap judgments, rather than a thorough analysis of the information at hand, resulting in errors or sub-optimal choices.

Financial institutions have established policies, procedures, and standards to identify and respond to cyber threats, but humans are always a weak point. Fraud attacks that exploit and capitalize on human trust – as AI-generated fraud often does – make end users a point of vulnerability to FIs.

Impostor Targets

- C-suite personnel
  - ➢ Purpose: To obtain access or sensitive information – such as company secrets, PII, and other non-public data – through impersonation.
  - ➢ Example: A deepfaked video impersonation of a C-suite executive sent to a lower-ranking employee enables a TA to initiate a transaction, transfer funds, or request access to nonpublic information.
- Banking firms
  - ➢ Purpose: Consumer fraud, money laundering
  - ➢ Example scenario 1: A criminal creates a website that impersonates a large FI and pays to ensure search page prominence to trick consumers into providing their credentials.
  - ➢ Example scenario 2: A criminal sends fraudulent text messages impersonating an FI to trick a customer into engaging with the criminal.
- Financial advisors/investment bankers
  - ➢ Purpose: Fraud

  - Example scenario: A deepfake video is used to impersonate financial advisors, investment bankers, or other trusted advisors to trick finance professionals into committing fraud.
- Third-party trusted relationships
  - Purpose: Fraud
  - Example scenario: Threat actors use deepfake impersonations of financial services staff or employees of their external partners to gain access to or exfiltrate money from financial services firms.
- Individuals/general public
  - Purpose: Espionage, revenue generation, sanctions avoidance, and initial access
  - Example scenario: Threat actors use GenAI to create fake identification documents to bypass human resource checks and gain employment for espionage, sanctions avoidance, or malicious activity.
- Public personas
  - Purpose: Embarrassment
  - Example scenario: A deepfake of a public persona is used to fool an FI’s C-suite into issuing a response that embarrasses them and discredits their security and vetting procedures.

Education & Awareness

AI-enabled fraud is evolving fast, but organizations do not need to wait for innovative technology to respond. Training, awareness, and robust controls offer an immediate practical line of defense as FIs explore longer-term technology solutions.

A critical component of a comprehensive fraud mitigation strategy is proactive education. According to the Association of Certified Fraud Examiners (ACFE), organizations with employee fraud training experienced a 47% reduction in fraud losses, while those with manager/executive training experienced a 50% reduction. One study indicates that with regular security awareness training, an organization can reduce phishing-related risks from 60% to 10% within the first 12 months.[6] Another study focusing on security awareness training reported a significant drop in phishing click rates, from 32% to just 5%, within a year.

Similar studies suggest that, as a sector, we can help change consumer behavior through practical fraud training, although the message requires regular updates due to diminishing returns over time.[7]

Furthermore, training develops employees' proactive risk mitigation skills. Trained employees are more vigilant and better able to identify potential threats, increasing the likelihood of timely incident reporting and mitigation.[8] For example, after implementing targeted security awareness training, a carrier company improved its phishing recognition rate by 90%, preventing a potential $2.6 million loss, and cyber education helped J.P. Morgan stop a $5.3 million fraud attempt.

The Goal of Fraud Education

Fraud education and awareness are fundamental to a financial institution's risk management framework. Education helps enable employees and customers to prevent, detect, and mitigate fraud. However, the speed, sophistication, and scale of AI-enhanced fraud demand an evolution in how FIs deliver education.

Unlike traditional fraud tactics, AI-enabled attacks can transfer funds in seconds, often making them irretrievable. This new landscape makes proactive, targeted education a strategic imperative.

FIs must structure effective AI-focused fraud education as a series of tailored campaigns to meet the needs of two core audiences: internal staff and external clients.

Each audience faces different threat vectors, necessitating customized messaging and training strategies grounded in relevance, repeatability, and clarity.

Internal Education Strategy

Audience Segmentation and Threat Alignment

Internal audiences must receive education about the fraud risks associated with their specific roles and communication channels. For example:

- **Call Center teams** must be trained to recognize AI-generated voice deepfakes and social engineering aimed at bypassing authentication protocols.

- **Underwriting departments** should be equipped to detect AI-generated documentation and fraudulent application materials.

Each financial institution's structure will vary, so education should begin with an audit of existing risk assessments to identify divisions, channels, and fraud exposure.

Example Risk Mapping Chart

Level of Risk	Deepfake	False Docs	Phishing	Account Compromise	BEC
Call Center (frontline)	High	Mid Low	High	High	Mid Low
Teller (frontline)	Low	High	High	Critical	Mid Low
System Support (frontline)	High	Mid Low	High	Critical	High
Loan Officer (frontline)	Critical	High	High	High	High
Underwriting (back office)	Medium	Critical	High	Medium	Medium
In Clearing (back office)	Medium	Critical	High	Mid High	Low
Funds Transfer -wires/ACH (back office)	Mid High	High	High	Critical	Critical
Cards - Debit and Credit (frontline)	High	Medium	High	High	High
Cards - Debit and Credit (back office)	Medium	High	High	High	Medium
Account Opening (back office)	High	Critical	High	Critical	Mid Low
Marketing	Mid Low	Mid Low	High	Low	Low
Fraud Operations	Critical	Critical	High	Critical	Critical
BSA/AML	Critical	Critical	High	Critical	Critical
Human Resources	Medium	Critical	High	Low	Low
InfoSec/SOC	Critical	Critical	High	Critical	Critical

This mapping forms the foundation for a focused education campaign by assigning risk levels (e.g., critical, high, medium, low) to various divisions. This process helps deliver education appropriate to the division's exposure.

Training Design Principles

- **Relevance:** Tailor messaging to each division's risk profile. For example, deepfake risk training for call centers focuses on voice, while training for lending officers relates to voice and video deepfakes.

- **Repeatability:** Vary the delivery method – combine computer-based education (CBEs) with live sessions and infographics for better retention and reinforcement.

- **Clarity:** Present a clear problem, the proposed mitigation, and a concise explanation of how the solution reduces the threat. Every training course should include an actionable takeaway.

The following scenarios illustrate how an FI's internal awareness/training strategies can be developed.

1. FI A noted an increase in loan defaults that it traced to fraudulent loan applications.
2. An internal investigation of the applications indicated a circumvention of FI A's training and awareness strategies: distributing examples of known fraudulent documents for team meeting discussions, a straightforward guide on some

external verification sources, and a video walking staff through effective verification processes.

3. FI A identified new strategies to validate documentation through external verification processes and cross-application document comparison.
4. FI A developed training focused on these strategies that detail the fraud mitigation steps and test the staff's knowledge. This allowed the FI to “step up " employee awareness through multiple media strategies for improved retention rates.

External Education Strategy

Client Segmentation & Threat Assessment

External audiences – consumers, small businesses, and corporate clients – face varied and often uncontrolled threat environments. Campaigns must avoid information overload by focusing on:

- The client’s primary communication channels
- Their digital banking behaviors
- Their specific fraud risk exposure

For instance, an FI focused on small business growth might prioritize education on AI-enhanced phishing campaigns targeting invoice or payroll systems. An institution aiming at consumer awareness may educate customers on deepfake “grandparent” scams.

Example Client Segmentation Chart

Account	% of total accounts	Growth
Consumer	45%	Not marketed
Small Business	20%	Primary target
Large Business	3%	Minimal marketing
Lawyer Accnt	2%	Minimal marketing
Managed Accounts	7%	Marketed
Trust Accounts	0%	Not marketed
RE Loan/HELOC	12%	Marketed
Commercial Loan	10%	Marketed
Gov Acct	1%	Not marketed

Campaign Design Principles

- **Relevance:** Use data and trends to match messaging with real-world fraud tactics. For instance, clarify how AI-generated phishing emails often bypass traditional red flags, such as grammatical errors.

- **Repeatability:** Deliver messages consistently but vary the format. Some options include:

    - Retail signage and mobile app banners for consumers

    - Online banking messages and in-person relationship manager sessions for small businesses

- **Clarity:** Keep messaging concise. Offer links or QR codes that direct clients to more in-depth educational content for those who want additional details.

Below is an example of an external communication/awareness strategy.

1. FI B has seen an increase in deepfake attacks targeting business clients. These attacks use payees' voices and likenesses to redirect payments to the TA.
2. FI B determined that small business clients were at the greatest risk, as online banking was their primary banking method at FI B.
3. FI B developed a targeted awareness program for its small business clients.

a. A banner on the online banking page warned of the risk of deepfake attacks and directed the clients to verify suspicious changes to vendor requests through a callback or direct meeting confirmation.
b. A live webinar education opportunity for small business clients to discuss deepfakes and mitigation strategies.
c. A detailed white paper describing the risks of deepfakes, prevention measures, and response actions.

Example Delivery Channels & Communication Tactics

Audience	Primary Channels	Format Examples
Internal staff	Intranet, email, learning management system	CBE, infographics, live webinars
Consumers	Branch signage, app banners	Posters, mobile alerts, short videos
Small business clients	Online banking portal	Infographics, alerts, relationship meetings

Education is a force multiplier in the fight against AI-enabled fraud. Financial institutions can transform awareness into resilience by equipping internal staff with targeted knowledge and empowering customers to recognize and avoid manipulation.

Success depends on:

- **Clear segmentation** of audiences based on their role or relationship with the institution.

- **Tailored and timely content** that resonates with their real-world behaviors.

- **Delivery methods** that balance cadence with cognitive load.

Fostering a culture of vigilance makes institutions more agile and adaptive – essential qualities in defending against the next generation of AI-driven fraud.

Incident Response & Operational Reporting

Attacks will arrive faster, target more victims simultaneously, and evolve in real-time, so FIs cannot rely solely on today's response models. Financial institutions must develop innovative, non-traditional approaches to combat these AI-driven attacks.

One aspect of that is aligning the full spectrum of fraud management functions (i.e., fraud strategy, fraud prevention, incident response, and intelligence teams), along with other critical business and security functions (i.e., product and business, technology, cyber, and risk), and adequately resourcing them to act quickly. Further, financial institutions require effective cross-organizational integration of customer service, cyber, and fraud teams to respond promptly to threats.

To manage the scale and velocity of AI-enhanced fraud, financial institutions must prioritize internal planning and readiness. Just as cybersecurity teams regularly conduct red-team or tabletop exercises, fraud units should run simulations of AI-enabled attacks to test escalation paths, decision-making authority, and cross-team coordination.

By practicing how to handle a sudden wave of highly targeted, AI-driven scams, institutions can reduce reaction time, identify gaps, and strengthen resilience against attacks that may unfold at a scale well beyond what they encounter today. The ability to manage these incidents will become a critical aspect of how well FIs mitigate attacks.

When a fraudulent activity or scam occurs, FIs can take various measures to improve the ecosystem.

- Sharing TTPs (tactics, techniques, and procedures) to help other institutions identify or prevent fraud.
- Using intel feeds, such as FS-ISAC's Cybera mule account feed, to help financial institutions stop a fraud attempt in progress.
- Sharing with law enforcement to help lead to actions against the fraudsters.
- Sharing information internally to help staff identify threats and/or train analytics teams to detect and prevent them.
- Obtaining insight into known scams and fraud to provide a better basis for consumer and staff education.

Cost Considerations

AI-generated fraud makes budgetary issues more complex. For instance, FIs often have to consider how:

- Investments in fusion centers, technology, consumer education, and staff training affect fraud and scam losses
- The internal use of AI to detect fraud and scams impacts costs
- Smart friction that enables consumers to identify more scams slows transactions

Trend data on investments and losses to the financial institution or consumer will provide decision-makers with information on the return on investment for activities to detect and prevent fraud and scams.

Historically, threat actors have targeted bigger financial institutions due to their larger digital footprint and more lucrative payoff relative to the time required to orchestrate an attack. However, given the economies of scale enabled by AI, the shift to targeting smaller institutions more frequently has become more attractive to malicious actors.

A rise in attack frequency would increase the volume of incidents an organization would need to mitigate through its incident response program. Nonetheless, an appropriate response to a surge of AI-assisted attacks would not demand a drastic change to current operations, but rather relatively incremental enhancements to the existing processes in place.

Indeed, organizations that have established a successful incident response program do not need to rebuild from scratch; rather, they can enhance their current capabilities and become more agile to respond appropriately to the ever-changing TTPs of fraudsters. Importantly, they can prioritize data reporting of cyber incidents and share relevant information among financial sector peers to enhance each organization's preparedness.

Defenders – including third-party providers – are increasingly using the same tools TAs leverage to combat fraud across multiple organizations simultaneously. Signature-based detection models remain a core component of any comprehensive security strategy; however, an innovative approach is necessary to detect both known and unknown threats effectively. The development of AI-based incident response platforms increases the speed at which the ecosystem can analyze incidents, assess threat severity, and estimate potential damage, thereby accelerating response and mitigation efforts. Although introducing AI platforms within an organization presents its own set of risks, technological advancements such as these could help teams mitigate fraud more quickly and with greater accuracy.

Deepfake Response

While many FIs may not have specific use cases for deepfakes, their containment strategies for many incidents already exist in their incident response procedures. Indeed, common sense and a healthy dose of skepticism are key: the simplest defense against voice or video deepfakes is to insist on calling back on a known good number. Fraudsters will likely end the interaction.

However, customers and employees can also benefit from awareness of deepfake-specific indicators, such as distorted or blurred faces, out-of-place objects, or strange backgrounds in video conferences. Asking a question only the real person could answer can uncover a fraud, as can the request for the performance of a simple task, like a hand wave, which the deepfake tool may poorly replicate.

With any attack, and especially with deepfakes, containment strategies are essential. The recommendation is to, whenever possible, limit the blast radius of any attack to reduce the damage if an incident has occurred or is believed to have occurred. At such times, FIs should:

- Escalate appropriately through their notification tree.
- Notify law enforcement, when appropriate, to activate financial fraud kill chains and prevent further escalation.
- Initiate communication between fraud units and cyber teams.

Deepfake Use Cases

- Finance worker pays out $25 million after video call with deepfake ‘chief financial officer’ | CNN
- Deepfake Scams Are Distorting Reality Itself | WIRED
- In 2024, a retiree in New Zealand lost around $133,000 to a cryptocurrency investment scam after seeing a Facebook advertisement featuring a deepfake of the country’s prime minister encouraging people to buy in.

Controls & Technology Response

Many of the controls necessary to counter AI-enabled fraud are already in place within FIs today, including authentication protocols, transaction monitoring, customer education, incident response, and risk assessments. Training and monitoring are inherent in those controls, and remain the backbone of controls to combat fraud – the voice and visual checks built into financial services controls to verify customers are one example.

Nonetheless, AI can help criminals create more convincing fake identification and more effective attacks. KYC controls are critical, but may not be sufficient. Education needs to shift to inform employees and customers about what the criminals do *and* what financial institutions will never do (e.g., ask customers to download software). Risk assessments must be updated to identify AI-driven fraud and scams.

To meet the increased global demand for fraud detection technology, driven in part by more stringent global regulations and the liability of financial institutions, the ecosystem is developing innovative technologies to address the need for cyber-enabled fraud detection and prevention. Yet as financial services companies find ways to prevent fraudsters, the fraudsters will adapt and modify their methods of attack. The risk will evolve as technology and processes advance.

The speed at which TAs can create AI-powered fraud attacks has changed the requirements for combating fraud with fraud intelligence. Adversaries can pivot at an astonishing rate and change their TTPs. Innovation is essential to expedite the delivery of fraud intelligence and implement the necessary changes to detect and prevent fraud.

Representational Outcomes for Consideration

AI increases the speed, scale, and realism of fraud attempts, meaning FIs must evaluate controls and stress-test for effectiveness in this unfamiliar environment. The challenge lies in ensuring they deliver the proper outcomes against the evolving threat. Institutions should therefore assess whether current measures are calibrated to detect and disrupt AI-driven schemes, adapt them as necessary, and embed continuous improvement to ensure fraud defenses remain resilient as the threat landscape evolves. FIs should also consider where augmenting with nascent technology would be additive.

The following examples illustrate practical ways financial institutions can adapt existing fraud controls to the realities of AI-enhanced scams. Each outcome represents an actionable measure that builds on proven fraud-management practices rather than replacing them.

1. Understand the Latest Fraud Behaviors, TTPs, and Anomalies

FIs need to understand customer baselines and identify when transactions deviate from the norm or replicate known fraudulent patterns. Some systems can analyze historical patterns across transaction amounts, timing, geolocation, device usage, and transaction frequency. When a customer deviates significantly (e.g., by transferring unusually large sums, using a new device from an unfamiliar location, or making multiple transfers to unfamiliar accounts), the system may flag the transaction for review or automated intervention.

> Use Case: Detecting a sudden, international $5,000 wire initiated by an elderly customer who usually makes small in-town purchases.

2. Monitoring Transactions in Real-Time for Scams

With a baseline of known-good and known-bad activity, institutions can implement systems that assess real-time transactions against indicators associated with scam activity. These may include the use of language in payment descriptions (e.g., “urgent,” “emergency,” “gift”), the timing of transactions, and the sudden addition of new recipients or unusual routing. Nascent rule-based and machine-learning-driven systems excel at these activities.

> Use Case: Flagging a wire transfer labeled "bail payment" initiated late at night to a newly added recipient overseas.

3. Deepfake Detection

As scammers become more adept at voice cloning and video deepfakes, FIs should consider integrating deepfake detection tools into their customer interaction channels. These tools use AI to analyze real-time voice cadence, lip-sync mismatches, or digital manipulation artifacts.

> Use Case: Intercepting a fraudulent call to a fraud hotline where a cloned voice of a CEO is trying to approve an unauthorized transaction.

4. In-App Warning Systems and Intelligent Friction

An FI's digital banking interfaces should include dynamic, context-aware warnings that appear when risky behaviors are detected. Adding friction – like requiring a call with a representative – can disrupt the scam process without overly inconveniencing the user. For instance, if a customer is about to complete a high-risk transaction, the app could display a targeted warning: "Are you being asked to send money by someone on the phone or online? This may be a scam."

> Use Case: A pop-up alert prevents the customer from completing a payment to a fake "tech support" company while the customer is still on the phone with a scammer.

5. Means for Secure Messaging Between Customers and the FI

Phishing and spoofing via SMS and email remain common vectors for scams. FIs could prioritize secure messaging, such as in-app messaging, for all sensitive communications and discourage reliance on SMS or email, which is easily spoofed.

> Use Case: A customer confused about an urgent request from the FI receives confirmation through the FI's app that no such request was made, potentially stopping a scam in progress.

6. Embrace Enhanced Authentication Strategies

As criminals increasingly use AI to create convincing phishing and deepfake scenarios, traditional multifactor authentication (MFA) methods, such as one-time passwords or SMS codes, are no longer sufficient. TAs can easily intercept these or socially engineer their discovery. Instead, financial institutions should prioritize phishing-resistant authentication, such as FIDO2 passkeys, which use cryptographic verification bound to the customer's device and cannot be reused or stolen through spoofed websites. Passkeys reduce reliance on passwords and SMS-based codes, preventing many credential-harvesting attacks that AI now makes more believable and scalable.

> Use Case: A customer receives a deepfake text or email urging them to log in through a spoofed FI link. When they attempt to do so, the passkey does not work because the fake site lacks the legitimate cryptographic relationship to the customer's device. The customer's account remains secure, and the institution can log and analyze the failed phishing attempt to improve detection intelligence.

7. Teaching Customers to Detect Fraud

Enabling customers to understand and help reduce fraud is a necessary part of the solution. FIs can offer customers embedded tools allowing them to screen messages, emails, or websites for scam indicators. These tools can leverage generative AI detection models trained to recognize phishing patterns or deepfake content. Usage can also provide feedback that enhances ecosystem tools. FIs should enable customers to customize risk controls on their accounts, such as the speed of payments or the types of payments supported online.

> Use Case: An FI deploys capabilities in its mobile app to help detect scam messages or phone calls.

8. Allowing for a Human-in-the-Loop for Escalation

FIs need the ability to verify intent and educate customers, especially when they detect high-risk transactions. Mandatory human intervention would be constructive, such as a video verification or a phone call with a fraud specialist.

> Use Case: Before processing a flagged transaction, the customer must complete a live video verification call, during which a trained fraud analyst confirms the legitimacy of the transfer.

9. Ability to Integrate Cross-Channel Scam Intelligence

Scams often involve multiple channels, including social media, text messages, phone calls, and online banking interfaces. Institutions should unify data streams across these platforms and apply natural language processing and entity recognition to identify emerging scam narratives and attack vectors.

> Use Case: Detecting that multiple customers are receiving messages referencing a fake contest linked to the FI and proactively warning all users.

10. Participation in Industry-Wide Threat Intelligence Networks

Fraudsters often reuse tactics and move between FIs. Sharing scam indicators – such as suspicious account numbers, fraud narratives, and known mule accounts – can help disrupt scams across the ecosystem.

Use Case: An FI identifies a scam and shares the scammer's known recipient account details with other FIs, preventing repeat offenses.

Collaborative Efforts to Prevent & Disrupt Fraud

Given the speed and sophistication of AI-enhanced fraud, actionable fraud intelligence sharing is a critical control. Criminals share and collaborate often and well. FIs need to do even better, engaging within the financial sector and across adjacent sectors. There has been some progress, but there is much more room for improvement, including cross-sector collaboration with telco and messaging providers.

Significant barriers to effective sharing are the lack of agreement within the ecosystem on shareable attributes, which would enable proactive awareness of emerging TTPs, and time limitations.

Conversely, efficient, straightforward processes for sending and receiving intelligence would enable more effective transmission, and clear guidance from regulators would promote executive sponsorship for sharing.

Common Obstacles to Sharing

- Regulatory sharing clarity
- Resource limitations
- Executive buy-in

Threat actors rarely target a single institution. They maximize their techniques by hitting multiple institutions and often use the same resources to facilitate their scheme. When shared across the sector, actionable intelligence on malicious activity leads to prevention for all.

A selection of the known sharing initiatives and frameworks relevant to fraud in the financial services sector includes the following.

Known Intelligence Sharing Initiatives and Frameworks	
FS-ISAC Cyber Fraud Prevention Framework	A framework to define the fraud lifecycle from cyber entry points to fraud monetization. Provides a protocol for partnership on fraud response, a common structure and lexicon to identify knowledge gaps, and a method for sharing indicators and lessons learned with the financial sector.

Fraud intelligence exchange communities	Examples: ABA, EWS, FS-ISAC, NCFTA, NEFCC, vendors
BPI-BITS Messaging Intelligence Sharing Pilot	Currently in the proof-of-concept stage, the initiative includes telecommunications and messaging providers.
Aspen FSP National Task Force for Fraud & Scam Prevention	Includes key cross-sector organizations in a united effort.
Threat Abuse Escalation Programs	Programs that allow a direct path for members of the financial sector to report known bad activity to partner organizations.
GASA Global Signal Exchange	The threat of AI-driven fraud and scams extends far beyond the United States. Recognizing its global impact, industry leaders have formed the Global Anti-Scam Alliance (GASA), uniting representatives from around the world. GASA brings together law enforcement agencies, government bodies, consumer protection organizations, financial authorities, brand-protection experts, social media platforms, internet service providers, and cybersecurity firms to share intelligence and coordinate action. This coalition reflects a strong, coordinated response to both individual and organized criminal networks. By breaking down traditional barriers to information sharing and collaboration, GASA aims to accelerate joint efforts that reduce consumer risk and enhance financial safety worldwide.

Other sectors provide inspiration and opportunities for collaboration. An emerging area of collaboration is with the AI technology providers. These platforms play a critical role in ensuring their technologies are not weaponized to enable large-scale impersonation of FI brands, executives, or customers.

As tools like OpenAI, ElevenLabs, and others become widely available and increasingly sophisticated, they carry the potential for innovation – as well as the risk of misuse. To that end, AI developers should implement proactive processes to detect and disrupt fraudulent use of their systems, including monitoring for impersonation, flagging synthetic content that mimics financial institutions, and embedding safeguards into model outputs.

Equally important is direct engagement with financial institutions. FIs cannot address AI-enabled fraud risks in isolation; they need reliable channels to notify AI firms of brand abuse and receive timely support in mitigating these risks. By establishing formal collaboration mechanisms, such as data-sharing arrangements, dedicated abuse response teams, or industry-wide safe harbor frameworks, AI platforms and financial institutions can jointly mitigate the risk of AI-enhanced fraud before it reaches consumers. This type of partnership will become increasingly essential as generative tools expand in capability and accessibility. FIs should prioritize these partnerships alongside existing collaborations with telecom providers, social media companies, and regulators.

The Need for Tracking AI-Enabled Scam Claims

As scams increasingly leverage generative AI, financial institutions and regulators must enhance their ability to detect and classify AI-driven scams. This is crucial for understanding evolving threat patterns, allocating resources effectively, designing targeted countermeasures, and meeting regulators' requirements.

Why It Matters

- **Trend analysis:** Knowing which scams are powered by AI (e.g., voice cloning, deepfake videos) allows institutions to identify which tools and narratives are gaining traction.

- **Resource allocation:** As AI-enabled scams rise sharply, more investment may be warranted to counter deepfake technologies and provide customer education.

- **Policy and insurance implications:** Insurers and regulators may require proof of AI involvement in fraud claims for coverage or enforcement purposes.

- **Regulatory implications:** Classifying attacks as AI-backed may be necessary to meet compliance with SARs and other regulatory requirements.

Challenges in Capturing This Data

- **Lack of customer awareness:** Victims may not realize they interacted with an AI-generated persona or a fake voice.

- **Absence of structured reporting:** Many scam databases and fraud case management systems lack a field for "AI involvement."

- **Attribution complexity:** Unlike phishing links or IP addresses, generative content leaves fewer technical fingerprints that can be definitively tied to AI usage.

- **Legal and ethical barriers:** Due to privacy concerns, some institutions may hesitate to analyze customer interactions deeply enough to determine AI involvement.

Recommendations

- Update fraud reporting taxonomies to include tags for suspected AI involvement.

- Train fraud analysts to recognize signs of AI-generated scams.

- Encourage customer reporting with simple prompts such as, “Did the person contacting you sound unusual or robotic?”

AI Companies Helping to Fight Fraud

As the adoption of artificial intelligence accelerates, the same tools that empower innovation are also being weaponized by criminals. Generative AI models, used to create realistic text, voices, and images, now enable fraudsters to impersonate financial institutions, replicate branding, and produce highly convincing deepfakes, dramatically increasing the success rate of scams. Many of these attacks are launched using commercial AI platforms, often without the provider’s awareness. This dynamic highlights a growing, urgent need for AI companies to play an active role in preventing fraud and protecting the integrity of their technologies.

AI companies must recognize that their platforms are part of the fraud ecosystem, both as potential enablers of attacks and as powerful tools for detection and defense. These firms should implement robust mechanisms to detect and prevent the misuse of their tools for criminal purposes, including the generation of synthetic media used in impersonation scams or the automated creation of fraudulent communications. Equally important, they should develop transparent reporting and response pathways for FIs and regulators to flag misuse, ensuring that confirmed abuse is swiftly contained and that the responsible accounts are terminated.

Collaboration is key. AI providers can help protect consumers and the financial sector by sharing indicators of impersonation activity, such as synthetic voice or image models trained on known financial brands, and by engaging directly with industry groups, including FS-ISAC, BPI, and ABA, to exchange threat intelligence. This partnership model would mirror long-standing collaborations between FIs, telecoms, and cybersecurity vendors, enabling the faster disruption of scams before they reach consumers.

Finally, AI companies should establish dedicated brand-protection programs to prevent their tools from being used to mimic or misappropriate the identities of financial institutions. These programs should include proactive blocking of FI names, logos, and likenesses in model outputs; brand-safety filters that detect attempted impersonation; and direct notification channels for FIs when brand misuse is detected. Such measures would significantly hinder criminals' ability to generate convincing deepfakes or spoofed materials that exploit consumer trust in well-known financial brands. By embedding brand integrity protections into their platforms, AI providers can materially reduce downstream fraud risk and demonstrate their commitment to responsible AI deployment.

Conclusion

The contributors to this paper believe that the existing controls are working and recommend maintaining and expanding FIs’ current controls rather than starting anew. Though AI is already increasing the volume of attacks, lowering entry barriers, and enhancing their accuracy and effectiveness, the types of attacks generally remain the same. We recommend making incremental enhancements to anti-fraud training and operations that incorporate AI concepts and best practices.

Just as adversaries are using AI to improve attack approaches, FIs can consider adopting a risk-based approach, including embracing nascent technologies, to enhance their existing defenses.

As this paper demonstrates, relying on common sense and lessons learned over time is a suitable initial approach to combating AI-enhanced fraud – but adversaries are using technology to circumvent and bypass these known strategies. By leveraging technology, the community can come together and identify new ways to protect itself and the sector from these enhanced attacks.

Adversaries can spend extraordinarily little on attacks, and defenders must spend much more to defend. FIs, telecommunications, media, and other sectors are in this fight. With proper education, incident response, controls, technology enhancements, and participation in large-scale efforts, we can mitigate the impact of these new attacks. This is a fight we can win.

References

R. Ramesh. Cloned Voice Tech is Coming for Bank Accounts. Bank Info Security. April 12, 2024. https://www.bankinfosecurity.com/cloned-voice-tech-coming-for-bank-accounts-a-24850

F. Heiding, B. Schneier, A. Vishwanath. AI will increase the Quality and Quantity of Phishing Schemes. Harvard Business Review. May 30, 2024. https://hbr.org/2024/05/ai-will-increase-the-quantity-and-quality-of-phishing-scams

S. Goswami. What are the Barriers to Cyber and Fraud Team Integration? Bank Info Security. May 14, 2024. https://www.bankinfosecurity.com/what-are-barriers-to-fraud-cyber-team-integration-a-25105

S. Goswami, R. Ramesh. What's Ailing Faster Payment Adoption in the US? Bank Info Security. April 3, 205. https://www.bankinfosecurity.com/whats-ailing-faster-payments-adoption-in-us-a-27918

OSFI-FCAC Risk Report - AI Uses and Risks at Federally Regulated Financial Institutions. September 24, 2024. https://www.osfi-bsif.gc.ca/en/about-osfi/reports-publications/osfi-fcac-risk-report-ai-uses-risks-federally-regulated-financial-institutions#toc-id-20

F. Natalucci, M. Qureshi, F. Sunthiem. Rising Cyber Threats Pose Serious Concerns for Financial Stability. April 9, 2024. https://www.imf.org/en/Blogs/Articles/2024/04/09/rising-cyber-threats-pose-serious-concerns-for-financial-stability

S. Sharma. AI in incident response: from smoke alarms to predictive intelligence. April 21, 2025. https://www.csoonline.com/article/3966034/ai-in-incident-response-from-smoke-alarms-to-predictive-intelligence.html

Dr. R. Poth. Utilizing AI for proactive threat hunting and incident response. October 18, 2024 https://rdene915.medium.com/utilizing-ai-for-proactive-threat-hunting-and-incident-response-4d63ad5bf442

I. Steinhaeuser. How AI will disrupt fraud detection and prevention technologies. Thomson Reuters. https://www.thomsonreuters.com/en-us/posts/corporates/technological-considerations-fraud-prevention/

L. Valk. 13 best fraud detection software solutions in 2024. Salv. https://salv.com/blog/fraud-detection-software-solutions/#verafin

V. Dave & A. Santhanagopalan. Google Cloud and Swift pioneer advanced AI and federated learning tech to help combat payment fraud. https://cloud.google.com/blog/products/identity-security/google-cloud-and-swift-pioneer-advanced-ai-and-federated-learning-tech?e=48754805

[1] FinCEN, *Suspicious Activity Report Filing Instructions: Deepfakes and Generative AI* (Jan. 2024)

[2] Europol, *Facing Reality? Law Enforcement and the Challenge of Deepfakes* (2022); Wall Street Journal, "Fraudsters Use AI to Scam Firms Out of Millions" (2023)

[3] Federal Trade Commission, *Consumer Sentinel Network Data Book 2023* (2024)

[4] University of California, Berkeley, *How Generative AI Improves Phishing Campaigns* (2023); Cybercrime Magazine, "Deepfake Audio Scams Cost Firms Millions" (2023)

[5] FinCEN Alert, *Deepfake and Synthetic Media in Identity Fraud* (2024)

[6] https://www.shrm.org/topics-tools/news/risk-management/fight-fraud-employee-awareness#:~:text=Fraud%20Awareness%20Training&text=Employee%20tips%20are%20the%20most,without%20such%20programs%20in%20place.

[7] https://link.springer.com/article/10.1007/s10610-024-09573-1

[8] https://gtreilly.com/nonprofits/fraud-awareness-training-and-strong-controls-help-nonprofits-fight-fraud#:~:text=The%202024%20ACFE%20report%20contains,losses%20of%20organizations%20with%20it.