Mitigating AI-Powered Attacks Against Identity and Authentication

Table of Contents

Executive Summary

It's a cautionary tale: in a well-publicized case, a Hong Kong finance worker received an email from his boss requesting a money transfer. After initially thinking it was a phishing

attempt, the worker joined a video call with his firm's chief financial officer and other team members to verify the funds transfer. Everything seemed normal, so he sent the money, only to later learn it was a deepfake attack that led to the loss of$25 million.[1]

The malicious use of generative AI (Gen AI) poses a threat to financial institutions' (FIs) identity proofing and authentication systems – in large part because Gen AI tools dramatically lower the barrier for creating convincing synthetic identities and sophisticated attacks.[2] Attackers are already using Gen AI to commit fraud through a variety of channels. Deepfakes – of videos, voices, and images – along with Gen-AI-composed phishing emails, can be virtually indistinguishable to human perception, creating more realistic, ever-evolving attack vectors. And these attacks are on the rise, with nearly 50% of companies experiencing some form of deepfake attack, up from 29% two-years ago.[3] While some of these attacks pose a direct threat to the identity proofing and authentication systems FIs use, others exploit social engineering in conjunction with deepfake technologies to attack FI systems.

Globally, 92% of businesses have experienced economic loss due to deepfakes, with an average cost of $450,000.[4] For FIs, however, that loss rises to $600,000. In the U.S., deepfake incidents in the FI sector increased by 700% in 2023 compared to the previous year.[5] Deloitte's Center for Financial Services predicts that Gen AI could enable fraud losses in the U.S. to reach $40 billion by 2027, up from $12.3 billion in 2023, representing a compound annual growth rate of 32%.[6] Meanwhile, Gen-AI-powered language models have enabled attackers to craft highly personalized, convincing phishing messages at scale, making these scams harder to detect and more effective at convincing individuals to disclose sensitive information.[7]

Artificial intelligence is the predominant term for technologies that can learn, analyze, and predict based on data. Gen AI extends this capability and generates new content based on the data it receives.[8] In the context of identity, Gen AI — through deepfakes and phishing emails — poses the primary threat.

[1] https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk

[2] Generative AI is defined as the class of AI that emulates the structure and characteristics of input data in order to generate derived synthetic content. This can include images, videos, audio, text, and other digital content.

[3] https://regulaforensics.com/resources/deepfake-trends-2024-report/?success=true

[4] Ibid

[5] https://www.wsj.com/articles/deepfakes-are-coming-for-the-financial-sector-0c72d1e5

[6] https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2024/deepfake-banking-fraud-risk-on-the-rise.html

[7] https://www.weforum.org/stories/2025/01/how-ai-driven-fraud-challenges-the-global-economy-and-ways-to-combat-it/#:~:text=Now%2C%20with%20sophisticated%20AI%20tools,it%20take%20to%20fight%20back?

[8] AI Executive Oversight Group AI Lexicon

Not all applications of Gen AI are malicious when it comes to identity, however, both AI and Gen AI can also be leveraged as part of security tools to mitigate the threats posed by the same technologies.

This paper underscores the urgency for financial institutions to adopt advanced mitigation strategies to address these evolving threats. As these AI threats evolve, mitigations must evolve as well. There is no single solution for attack mitigation; rather, FIs need to pull a series of levers to protect institutions and consumers from these attacks. As GenAI advances, financial institutions face three primary attack vectors - comprising of 10 specific tactics that threaten their systems:

Attack Vector 1: Deepfake-Driven Social Engineering and Impersonation
AI-generated voices, videos, and messages impersonate individuals or organizations, amplifying social engineering attacks and making fraudulent interactions difficult to detect:

Tactic 1: Deepfake attacks on identity verification
Tactic 2: Social engineering using deepfake voices against a financial services call center or help desk
Tactic 3: Use of Gen AI to compose realistic phishing campaigns for credential compromise
Tactic 4: Targeting companies or individuals via AI-generated impersonation
Tactic 5: AI-generated deepfake biometric spoofing attacks to defeat authentication
Tactic 6: Cross-channel deepfake attacks

Attack Vector 2: Synthetic Identity Creation
GenAI is used to fabricate convincing identities by combining breached data with realistic imagery and documentation.

Tactic 7: Using Gen AI to create holistic, synthetic identities
Tactic 8: AI-generated synthetic identity fraud in remote hiring
Tactic 9: Real-time deepfake attacks used to enable synthetic identity fraud

Attack Vector 3: AI Agents as Attack Surrogates
Autonomous AI agents mimic user behaviour and bypass verification steps through automated, scripted interactions and high-volume transactions.

Tactic 10: Using AI agents for account takeovers and automated fraud campaigns

FIs need to prepare and use various resources and information-sharing systems to track threats, then prepare proper mitigations to disrupt and defend against them. This paper:

- Scopes the threat that malicious uses of Gen AI pose to identity systems for FIs.
- Details the specific threats Gen AI poses to financial services organizations.
- Describes mitigations to thwart the attacks and the considerations necessary for those mitigations.
- Previews emerging Gen AI threats to identity systems.
- Proposes a Gen AI Risk Mitigation Maturity Model for financial services organizations.

Purpose

The purpose of this paper is to highlight three current and emerging attack vectors powered by Gen AI – along with ten concrete examples of how adversaries are using these attack vectors to compromise the identity and authentication tools used by many FIs – and outline potential mitigations that FIs can deploy to guard against each of them.

By examining these threats and exploring how both GenAI and traditional AI can be leveraged for defense, the paper delivers practical insights to help financial institutions protect their systems and consumers from increasingly sophisticated fraud and identity risks.

The paper makes clear that there is no single solution for attack mitigation; instead, financial institutions can employ a range of approaches and leverage multiple resources and information-sharing systems to protect both institutions and consumers.

This paper is a deliverable of the Financial Services Sector Coordinating Council's Artificial Intelligence and Identity and Authentication Workstream (AI-IA).[9] Workstream participants include over 60 executives from financial services firms, technology companies and US government agencies. The American Bankers Association and Better Identity Coalition provided support to the FSSCC in developing this paper.[10]

Audience

This paper is intended for FIs, cybersecurity and fraud professionals, AI service providers, telecommunications companies, and policymakers at regulatory agencies and in legislative bodies who are responsible for safeguarding identity systems and mitigating the risks posed by Gen AI.

[9] The FSSCC is an industry-led, non-profit organization that coordinates critical infrastructure and homeland security activities within the financial services industry. Their members consist of financial trade associations, financial utilities, and the most critical financial firms. https://fsscc.org/

[10] The American Bankers Association is the voice of the nation's $25.1 trillion banking industry, which is composed of small, regional and large banks that together employ over 2 million people, safeguard $19.7 trillion in deposits and extend $13.2 trillion in loans. The Better Identity Coalition brings together leading companies to promote education, collaboration, and better solutions to protect identities online. Launched in early 2018, the coalition is an initiative of the Center for Cybersecurity Policy and Law, a nonprofit dedicated to working with policymakers to advance cybersecurity initiatives.

Overview

The following scenarios illustrate how attackers can exploit identity and authentication systems using Gen AI.

- Created via a Gen AI deepfake, an employee receives a video call from the CEO requesting the wiring of funds to an offshore consulting firm to help expand the business, but the call is not actually from the CEO.

- A customer service representative in a call center receives a phone call from someone claiming to be a customer seeking to transfer funds from a brokerage account and requesting a change in address, but it isn't the customer. Deepfake technology is spoofing the customer's voice.

- A consumer receives an email from their bank asking to confirm certain personal data. The customer clicks a link, enters their username and password, then enter a one-time passcode, only to receive an error code from the site because the customer was phished.

Each of these attack types has been around for years, but the availability of highly accessible, cheap, convincing Gen-AI tools is supercharging their potency and enabling adversaries to launch them at scale. Prior to the use of jailbroken Gen-AI models to develop convincing deepfakes, threat actors looking to spoof a CEO in a video call might have relied on a latex mask and poor lighting to deceive an employee into believing that they were the CEO. Today, due to the widespread commercialization and adoption of Gen-AI models in criminal communities, these attacks can be launched much more easily and convincingly using deepfake technology.

To be clear, not all these attacks are direct assaults on FI identity verification and authentication components. Fraudsters are combining Gen-AI technology with social engineering to commit fraud. That doesn't lessen the threat that jailbroken Gen-AI technologies pose to the identity systems that FIs rely on for identity proofing and authentication. These attacks vary, with some directly targeting FI identity proofing and authentication systems, while others attack the FI's consumers or employees with a combination of social engineering and deepfakes, leveraging audio, video, and photographic deepfake technologies. The common theme is that criminals can now use Gen AI to easily impersonate others to perpetrate fraud.

The risk to FIs is immense. More than 105,000 attacks leveraging AI-generated deepfakes were launched in the U.S. alone last year.[11] Using urgency and authority, scammers trick employees into wiring funds, sharing sensitive data, or clicking on malicious links. Deepfakes are harder to detect as tools improve and executives' public videos provide

[11] https://www.wsj.com/articles/ai-drives-rise-in-ceo-impersonator-scams-2bd675c4

ample training data. Losses topped $200 million in Q1 2025; however, incidents go unreported to avoid reputational damage.[12]

Employees and customers need to be vigilant when reviewing emails, as 60% of individuals have fallen victim to AI-automated phishing.[13] While the percentage is on par with other phishing attempts, AI enables fraudsters to conduct these attacks at scale, faster, and more cheaply. The entire phishing process can be automated using Large Language Models (LLMs), reducing the cost of phishing attacks by more than 95% while achieving equal or greater success rates.[14]

While phishing poses a persistent threat, deepfake attacks are also on the rise. Deepfake incidents in the fintech sector increased by 700% in 2023 compared to the previous year.[15]

Deloitte's Center for Financial Services predicts that Gen AI could drive fraud losses in the United States to $40 billion by 2027, up from $12.3 billion in 2023, representing a compound annual growth rate of 32%.[16]

Quantifying the Threat of Malicious Uses of Gen AI

Hundreds of billions of dollars are lost each year due to fraud and identity theft. Malicious use of Gen AI to augment the existing attacks may dramatically increase loss amounts. In January 2024, the Financial Crimes Enforcement Network (FinCEN) issued a Financial Trend Analysis on information linked to identity-related suspicious activity in Bank Secrecy Act reports filed in 2021 and found that 1.6 million, which represents 42 percent of Suspicious Activity Reports (SARs) filed that year, were tied to a compromise of identity or authentication. This translates to $212 billion worth of transactions.[17]

While FinCEN did not specify how many of these SARs involved AI-powered attacks, of the 14 typologies it created to categorize identity-related attacks, all but one can be exploited by AI.

FinCEN released a follow-up Alert in November 2024, which notified FIs of new fraud schemes involving Gen-AI-powered deepfakes.[18] While the alert did not quantify the impact of deepfakes on the SARs FinCEN received, it created a new code "Fin-2024-DEEPFAKEFRAUD" for FIs to use when filing SARs tied to a deepfake.

---

[12] https://www.wsj.com/articles/ai-drives-rise-in-ceo-impersonator-scams-2bd675c4

[13] https://hbr.org/2024/05/ai-will-increase-the-quantity-and-quality-of-phishing-scams

[14] Ibid

[15] https://www.wsj.com/articles/deepfakes-are-coming-for-the-financial-sector-0c72d1e5

[16] https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2024/deepfake-banking-fraud-risk-on-the-rise.html

[17] https://www.fincen.gov/news/news-releases/fincen-issues-analysis-identity-related-suspicious-activity

[18] https://www.fincen.gov/sites/default/files/shared/FinCEN-Alert-DeepFakes-Alert508FINAL.pdf

The alert also defined nine red flags designed to help financial institutions spot when illicit Gen-AI tools may be used. These flags include:

1. A customer's photo is inconsistent, shows visual signs of being altered, or is inconsistent with their other identifying information. For example, a customer's date of birth indicates they are much older or younger than their photo suggests.

2. A customer presents multiple forms of identity that are inconsistent with each other.

3. A customer uses a third-party webcam plugin—software that augments the camera's functionality—during a live verification check. Additionally, a customer attempts to change communication methods during a live verification check due to excessive or suspicious technological glitches during remote verification of their identity.

4. A customer declines to use multifactor authentication (MFA) to verify their identity.

5. A reverse-image lookup or open-source search of an identity photo matches an image in an online gallery of AI-produced faces.

6. A customer's photo or video is flagged by commercial or open-source deepfake-detection software.

7. Gen-AI detection software flags the potential use of Gen-AI text in a customer's profile or responses to prompts.

8. A customer's geographic or device data is inconsistent with the customer's identity documents.

9. A newly opened account or an account with little prior transaction history has a pattern of rapid transactions; high payment volumes to potentially risky payees, such as gambling websites or digital asset exchanges; or high volumes of chargebacks or rejected payments.

Defining the Threats and Mitigations

Gen AI can take video, voice clips, and create a realistic model of an individual that can be superimposed over another person to impersonate them. The use of these tools to attack identity is on the rise. In 2024, a deepfake attack happened every 5 minutes, while digital document forgeries increased 244% year over year.[19] With online identity verification playing a vital part in the financial services onboarding process, detecting these attacks in real time is critical to preventing fraud and financial crime.

But attackers are not just focused on compromising tools used to onboard new customers. Phishing attacks focused on stealing login credentials are increasing exponentially. In the second half of 2024, credential phishing soared by 703%, signaling a sharp escalation in the use of sophisticated phishing kits and social engineering tactics.[20]

Below, we outline the three primary Gen-AI-enabled attack vectors along with ten common tactics used by adversaries to compromise identity and authentication systems, along with potential mitigations for each attack and tactic.

Attack Vector 1: Deepfake-Driven Social Engineering and Impersonation

AI-generated voices, videos, and messages impersonate individuals or organizations, amplifying social engineering attacks and making fraudulent interactions difficult to detect.

Tactic No. 1: Deepfake Attacks on Identity Verification

> "Alice" is opening a new credit card account on a bank's website. She enters her demographic information, Social Security number, and a photo of her driver's license, and submits a selfie to verify that she is the same person to whom the license was issued. She passes the test, and the credit application is approved.
>
> The problem is that Alice used a stolen identity paired with a deepfake driver's license to fool the identity verification system. Fraudsters can purchase deepfake images of driver's licenses with the fraudsters picture on them – paired with real data from a stolen identity – to make it easier to fool identity verification systems.[21]

Mitigation: Using an image analyzer tool to look for photo editing and to make sure that necessary security features are present on the document. Liveness detection tools—which can help detect spoofed IDs or photos—should also be in place.

---

[19] https://www.entrust.com/resources/reports/identity-fraud-report?utm_source=press&utm_medium=onfido&utm_content=fraud-report&utm_campaign=ww-2025-fraudreport-onfidopressrelease-report

[20] https://slashnext.com/press-release/2024-eoy-phishing-intelligence-report/

[21] https://www.404media.co/inside-the-underground-site-where-ai-neural-networks-churns-out-fake-ids-onlyfake/

There are two types of liveness detection technologies: those that guard against "presentation attacks," which look to use a physical replica of a biometric, such as a photo, mask, or fake fingerprint, to trick a biometric system, and "injection attacks," which look to bypass the camera or biometric sensor completely to inject a fake image into the system. Of the two, injection attacks are used in deepfake attacks, and thus, liveness-detection technology that can detect and block them is quickly becoming the more important of the two. These systems focus on identifying attempts to bypass defenses by directly injecting manipulated media rather than capturing them through the intended camera or sensor. They aim to verify the authenticity of the data source and channel, preventing prerecorded deepfake files from being disguised as live image or video captures.

In addition, liveness detection tools may be either passive or active. Passive liveness detection tools analyze an individual's normal movements, such as blinking, slight head moves, and other natural reactions, and use those signals to determine liveness. Active liveness detection requires an individual to perform specific activities, such as smiling, turning their head to the right, and additional movements. FIs have found that consumers prefer passive liveness because it imposes less of a burden. FIs report seeing more customers abandon a transaction with active liveness detection.

The demographic data on identity documents should also be verified and validated by a mix of first- and third-party data sources; first-party sources might include government services such as the Social Security Administration's electronic Consent-based Social Security Verification (eCBSV)[22] or the American Association of Motor Vehicle Administrators Driver's License Data Validation (DLDV)[23] service. Third-party sources might include credit bureaus or security vendors that aggregate and analyze data to predict whether an identity is authentic or fraudulent, or a consortium-based network, such as the Financial Services Information Sharing and Analysis Center (FS-ISAC), which can notify about fraud and other emerging attacks. Additionally, risk-based technologies should be used to analyze device information, IP addresses, and other attributes to detect potential fraudulent behavior.

Institutions can also use commercial or open-source deepfake-detection software to spot attacks and deploy best practices and compensating controls from the Digital Identity Guidelines (NIST SP 800-63-4), which includes guidance on protecting against Presentation Attack Detection, injection detection, and other attacks.

Additionally, as mobile driver licenses (mDLs), verifiable credentials, and other credentials emerge, FIs can rely on them as roots of trust for identity verification. Institutions should also consider integrating AI-driven anomaly detection systems to flag unusual patterns in identity verification attempts.

[22] https://www.ssa.gov/dataexchange/eCBSV/

[23] https://www.aamva.org/technology/systems/verification-systems

FinCEN Flags relevant to this threat: 1, 2, 3, 5, 6, 7, 8.

Tactic No. 2: Social Engineering Using Deepfake Voice Against a Financial Services Call Center or Help Desk

> Brian, a customer service representative at a financial institution, works at a call center. The financial institution has implemented voice biometrics to help authenticate customers when they dial in by matching their voices against existing templates.
>
> "Roy" is calling in today because he needs to transfer $50,000 from a business account to the account of a new consultant to help with sales in another country. Before Roy even talks to Brian, he provides his account information and is authenticated via the voice biometrics system.
>
> Brian asks Roy standard questions concerning his Social Security number, date of birth, and address, and Roy confirms the identity information. The money is wired from the account, and the transaction is complete.
>
> While the bank used voice biometrics to secure accounts, deepfake technology spoofed Roy's voice and bypassed the system, and used other fraud techniques to steal account information to get past the customer service representative.

Mitigation: Ensure that call centers deploy multiple risk-based technologies to verify the "liveness" of a voice, device information, mobile device subscriber information, and other attributes from the individual calling in. Train employees to recognize subtle changes in voice timbre, cadence, or other oddities that may indicate the use of a deepfake. Have step-up processes in place that require additional steps when transferring funds, such as calling individuals back on another phone line and requiring out-of-band verification using a secondary trusted channel. Technology is emerging in this space to also enable video calls that address concerns about voice spoofing, though, as we discuss elsewhere in this paper, videos are also under attack from deepfakes.

FinCEN Flags relevant to this threat: 4, 8, 9.

Tactic No. 3: Use of Generative AI to Compose Realistic Phishing Campaigns for Credential Compromise

> Robert checks his email and sees an email from his bank with familiar graphics and messaging. The email notifies him of a change in his credit score due to a new account opening and asks him to click a link for further information.
>
> He clicks on the link and enters his username, password, and one-time passcode from an SMS notification. The page errors out, and after a couple of more attempts, he closes the browser and goes to a fresh page to log in, but finds that he is now locked out of his account; an attacker phished his credentials and then quickly changed the password, so he cannot get back into it.

> The attacker then transfers funds to other accounts.

Mitigation: Enable phishing-resistant authentication, such as passkeys, which can prevent credential compromise via phishing.[24] Even if an individual clicks on a link, the authentication data sent would not enable access.

In addition, emerging AI and machine learning tools can collect and process diverse threat intelligence data to predict and prevent attacks, as well as detect active threats.[25] For example, AI might analyze historical and ongoing incidents across a variety of organizations, types of cyberattacks, targeted geographic regions, organizational sectors, departments, and types of employees. The Open ID Foundation's Shared Signals Framework provides a valuable way for risk and authentication signals and events to be shared between cooperating peers.[26]

With this information, AI can help identify which types of attacks a given organization is most likely to experience and then automatically train security tools. Systems can also be used to detect whether funds are being transferred to accounts previously identified as risky or fraudulent. FIs can also place a human in the loop to decide on questionable, high-risk transactions.

Explore other alternatives, like intelligence sharing through the FS-ISAC and the Cybersecurity Infrastructure Security Agency's (CISA) Joint Cyber Defense Collaborative. FIs should also consider phishing simulation training and layered security measures such as Domain-based Message Authentication, Reporting & Conformance, Sender Policy Framework, DomainKeys Identified Mail, anti-malware and phishing detection, email encryption, and data loss prevention technologies. Enterprise and high-risk users should also use S/MIME signing and encryption, and update phishing simulation training to include scenarios involving AI-generated content to ensure employees are prepared for evolving threats. This could even include targeted phishing exercises based on function and location, using information gathered from social media.

FinCEN Flags relevant to this threat: 8, 9.

Tactic No. 4: Targeting Companies or Individuals via AI-Generated Impersonation

> Kevin's phone rings, and the caller ID says it's the chief financial officer (CFO) from his employer. Chris, the CFO, is a gregarious individual who always asks how Kevin is doing before jumping into details of the call. This call, however, is different. Chris says, "I need

[24] NIST has defined phishing-resistance in Section 3.2.5 of NIST SP 800-63B-4 at https://pages.nist.gov/800-63-4/sp800-63b.html#verifimpers. For more information on passkeys as an example of phishing-resistant authentication, see https://fidoalliance.org/passkeys/

[25] https://www.techtarget.com/searchsecurity/tip/Generative-AI-is-making-phishing-attacks-more-dangerous

[26] See https://openid.net/specs/openid-sharedsignals-framework-1_0.html

> you to transfer $300,000 immediately to Off-Shore Consultants LLP! They're going to help our expansion in the Asia Pacific."
>
> Kevin tries to ask follow-up questions, but Chris cuts him off. "Just send the money now. I emailed the wire information," Chris adds. Kevin checks his inbox and sees the information from Chris. "I'll wait on the line while you send the funds," Chris says.
>
> "It's done, I emailed the confirmation to you," Kevin responds. Chris thanks him and hangs up.
>
> Later that morning, Chris walks up to Kevin's desk and asks him about this wire transfer. "I didn't call you authorizing this," she says.

Mitigation: Training employees and consumers about this type of fraud is the first step to stopping it. Employees and consumers need to be trained to spot potential changes in voice timbre, cadence, video quality, or other oddities. They also need to be taught that fraudsters will use tactics like urgency to provoke a reaction and slow down the process.

These attacks are becoming increasingly popular and are not just against corporations.[27] Incidents involving deepfake phishing and fraud have skyrocketed by 3,000% since 2022, with a deepfake attempt occurring every five minutes in 2024.[28] Individuals are getting phone calls from family members saying they're in jail or have been kidnapped and need money. To address this risk, security experts are now recommending that families set up a passphrase they can use in an emergency to quickly determine if someone claiming to be a loved one is an impostor.[29] It is also recommended that families recognize the signs of these frauds, hang up on the caller, and call the person they know directly.

It is imperative to have processes in place that require additional steps when transferring funds, including calling individuals back from another line or requesting additional verification; implementing voice authentication passphrases for sensitive transactions; requiring OOB verification using a secondary trusted channel; ensuring call-back policies are in place, and other mitigations.[30]

FinCEN Flags relevant to this threat: 1, 3, 4, 6, 7.

Tactic No. 5: Deepfakes Against Authentication

> Christine is a high-net-worth individual. Fraudsters decide to target her bank accounts by using the device's native face verification to authenticate on the banking app.

[27] https://www.weforum.org/stories/2025/02/deepfake-ai-cybercrime-arup/

[28] https://www.entrust.com/company/newsroom/deepfake-attacks-strike-every-five-minutes-amid-244-surge-in-digital-document-forgeries

[29] https://vimeo.com/918408198?share=copy.

[30] https://www.sans.org/newsletters/ouch/beware-deepfakes-new-age-of-deception

> They surveil Christine and collect photos and videos from her online presence, including a YouTube video of her speaking at a conference last year. They steal her device and use data and AI-powered software to create a deepfake video to bypass security, gain access to her accounts, and transfer funds.

Mitigation: Use liveness detection to detect the possibility of deepfakes, and leverage risk-based technologies to verify device information, IP addresses, and other attributes to determine whether they are suspicious or may be associated with potential fraudsters. Use a consortium-based verification network, such as FS-ISAC, to receive notifications of new attack patterns. For high-risk transactions, FIs can use another authentication method, including a dynamic code, a passkey, email, text, or a push-based app. The bank may also require document verification, like what is used in the initial account onboarding process, a video call, or the individual physically appearing at a branch to conclude the transaction.

FinCEN Flags relevant to this threat: 1, 2, 3, 4, 6, 7, 8.

Tactic No. 6: Cross-Channel Deepfake Attacks

> Susan received a text message from her teenage son saying he's been in a car accident. She barely has time to respond before her phone rings with a call from her son. She answers.
>
> The individual on the other end of the phone sounds like her son. He's been injured in a car accident, and the police are taking him into custody because of it. He says he needs money immediately to avoid going to jail.
>
> He hangs up abruptly, and a text message appears with a link to where she can send money to keep her son out of jail.

Mitigation: Banks, telecommunications providers, and the government need to educate consumers about the threats posed by deepfake attacks and the potential for their use against friends and family, and to help prevent them, including setting up a shared password or passphrase to verify that a call is real.[31] FIs can launch educational campaigns for consumers and share information about these threats through methods such as ad campaigns, emails, and webinars.

[31] www.aba.com/Seniors

Attack Vector 2: Synthetic Identity Creation

GenAI is used to fabricate convincing identities by combining breached data with realistic imagery and documentation.

Tactic No. 7: Using Gen AI to Create Synthetic Identities

> "James Martin" is opening a new demand deposit account and applying for a credit card. He uploads his driver's license and demographic data, including his Social Security number, to start the process. The photo on the license is validated by an ID verification product, which also confirms that the photo on the card matches a selfie image James is asked to take. Little to no credit history is found, but nothing suspicious is discovered, so the account is opened, and a credit card is issued.
>
> Little does the bank know, James is a synthetic identity. A Gen-AI attack tool used information from data breaches to fabricate a new identity, then used deepfake technology to create the digital image of the driver's license and photos. The fraudsters used the same AI agent to open 100 accounts at various FIs across the country.

Mitigation: FIs need to layer multiple risk-based technologies to spot this type of fraud. Liveness detection, device signals, IP addresses, and other attributes could all be used to spot these types of synthetic identities. Likewise, checking the identity data against authoritative government sources such as eCBSV and the DLDV system can help to root out synthetic fraud. Government-issued verifiable credentials, such as mDLs, can play a crucial role in reducing the risk of deepfake attacks. If available, they can be used to cryptographically assert an identity. Institutions should collaborate with the public and private sectors to share intelligence on synthetic identity patterns and emerging attack techniques.

FinCEN Flags relevant to this threat: 1, 2, 3, 5, 6, 7, 8.

Tactic No. 8: AI-Generated Synthetic Identity Fraud in Remote Hiring

> Angela is hiring remote help desk technicians for her financial institution. She interviews candidates, and Jennifer is a finalist, but there are concerns. Jennifer has an unreliable internet connection and has used it as an excuse to rarely appear on camera during the interviews. Her resume is exceptional, and she has the necessary experience to hit the ground running from day one.
>
> When conducting some cursory social media background checks, there isn't much history, as most accounts seem to have been created recently. During interviews with team members, Jennifer has also given different answers regarding work history that conflict with her resume.
>
> When asked to complete a third-party remote identity verification process, Jennifer agrees. She provides the information to the third party. The information checks out as valid, as she's using a stolen identity and has used deepfake technology to create a fake

> driver's license whose data matches that identity. She receives an offer and begins work, but her laptop is set up in a North American laptop farm, which Jennifer remotes into from North Korea. From there, she probes the network and applications, then relays the information back to the North Korean government to use for future attacks.

Mitigation: Enabling proper training and processes will help FIs ensure they avoid hiring individuals from hostile nation-states. Organizations should consider compiling resource guides for security teams and recruiters to identify behavioral trends and red flags. There has been an increase in individuals applying for jobs from other countries to fund government programs, looking to cash in on a short-term paycheck, or steal intellectual property. Individuals are either using stolen or synthetic identities, having someone else go through the interview process, or using deepfake technology during the interview to look like someone else.

However, mitigations of this type of fraud are not all technology based. Interviewers should note inconsistencies in candidates' work experiences or the stories they tell. Not appearing on camera or being slow to respond to questions may indicate the subject is being coached.

That said, vendors offer AI-based fraud detection systems that can review an individual's IP address, device information, and other metrics to spot anomalous behavior and run it against databases of known fraudulent behavior. These same checks can be performed manually. Liveness detection can also be used during remote video interviews to spot deepfake technology. Additionally, companies should conduct thorough identity proofing of individuals, including document verification that uses liveness detection to ensure technology isn't being used. In some cases, companies should require individuals to appear in person to prove their identity and use a commercial service for in-person identity proofing.

The demographic data in identity documents should also be validated against third-party data sources to ensure it is genuine and not synthetic. Institutions should also verify the geographic location of remote employees during onboarding and periodically thereafter.

FinCEN Flags relevant to this threat: 1, 2, 3, 4, 5, 6, 7, 8.

Tactic No. 9: Real-Time Deepfake Fraud

> Mark connects with a woman online. "Sally" lives on the other side of the country, but they chat daily via phone. During a conversation, Sally reveals that her mother is sick and needs money for medications.
>
> Mark sends the funds. The conversations continue, and Mark sends Sally money to travel to see him. He's waiting at the airport to greet her, but she never shows up. When he tries to reach out, all the accounts she previously used are deactivated.

After reporting the incident, he finds out Sally never existed; she was a synthetic identity created with the help of deepfake photo and video technology, being operated by a perpetrator named Jack who resides outside the U.S.[32] He was part of a romance scheme that used real-time deepfake technology to pose as others and bilk them out of their savings.

Mitigation: Educate consumers about these threats and advise them to send money only to friends, family members, and individuals they have met in person. FIs, telecommunications providers, and the government can all play a role in helping educate the public. This technology has advanced significantly and will only get better, so the ability to spot it will be difficult. Having FinCEN, the FBI, CISA, or another agency track these threats and typologies and share information with FIs would also be helpful.

FinCEN Flags relevant to this threat: 1, 2, 5, 6, 7, 8.

Attack Vector 3: AI Agents as Attack Surrogates:

Autonomous AI agents mimic user behavior and bypass verification steps through automated, scripted interactions and high-volume transactions.

Tactic No. 10: Using AI Agents for Account Takeovers and Automated Fraud Campaigns

AI agents are emerging to help individuals manage mundane tasks and research. Looking for the best flight to Orlando in April? An AI agent can help. Researching the best mid-sized SUV? An AI agent can help with that, too.

But fraudsters could start using AI agents to perform account takeovers and organized fraud campaigns. They will involve scouring the dark web for usernames and passwords, and then using the internet to take over consumer accounts, potentially reducing the time to execute an attack by 50%, according to Gartner, Inc.[33]

AI agents will enable automation for more steps in account takeovers, from social engineering based on deepfake voices to end-to-end automation of user credential abuses.

AI shopping agents could also be tricked and scammed into purchasing from fake or deceptive storefronts. Such scam merchants may appear legitimate but are designed to defraud by manipulating the agent into purchasing from the scam site. This could raise challenges for FIs in validating the legitimacy of merchants and accurately authenticating the identities of AI agents with whom they share payment details or other sensitive information during transactions.

[32] https://consumer.ftc.gov/features/pass-it-on

[33] https://www.gartner.com/en/newsroom/press-releases/2025-03-18-gartner-predicts-ai-agents-will-reduce-the-time-it-takes-to-exploit-account-exposures-by-50-percent-by-2027

Mitigation: This paper has focused on the threat of Gen AI to FIs, but agentic AI is different. Agents will be able to act on behalf of individuals but may also help attackers. FIs need to deploy products, apps, APIs, and voice channels to detect, monitor, and classify interactions involving AI agents as attackers deploy them to attack systems and attempt to take over legitimate ones to commit fraud. The technology, standards, and protocols for authenticating and monitoring agents are evolving, and FIs need to watch the space and be aware of the latest developments.

FinCEN Flags relevant to this threat: 4,7, 8, 9.

The AI Agent Conundrum

Agentic AI refers to AI systems that can autonomously pursue complex goals, make decisions, and execute multi-step processes with minimal human intervention. It is characterized by its ability to act independently, adapt to changing conditions, and reason through tasks to achieve desired outcomes. The technology is designed to operate more like a human employee, exhibiting agency and taking initiative, as defined in the AI Executive Oversight Group AI Lexicon.

Agentic AI is a rapidly evolving field with limitless potential and limitless risk. These agents will be enabled to make transactions on behalf of individuals, but the standards and technical details of how this will all work are still in flux. FIs face myriad new challenges in addressing identity, authentication, and authorization requirements for agents, raising a new set of policy, regulatory, and standards issues that the government will need to help address.

The possibilities and challenges with this technology are truly endless from both the attack and mitigation perspectives. FIs will need to keep a vigilant watch on this technology to see how consumers use it to conduct transactions and how attackers exploit it to commit fraud.

Other Mitigation Considerations

In addition to the specific mitigation steps outlined above, FIs should regularly evaluate and validate their AI-driven defenses to ensure they remain effective, compliant with regulatory expectations, and carefully implemented to balance effectiveness, user experience, and privacy considerations. Other mitigation considerations include using AI to fight AI, user experience matters, regulatory and privacy considerations, scalability and deployment challenges, and policy.

Using AI to Fight AI

By leveraging AI to enhance security measures, financial institutions can better protect themselves and their customers from the evolving threats posed by Gen AI. This proactive approach helps maintain trust and ensures the integrity of identity systems.

- AI-driven liveness checks can help distinguish real users from deepfakes, replay attacks, injection attacks, or manipulated videos. This technology is often coupled with document verification tools in which the ID proofing product cross-checks a selfie photo against the photo on a government-issued ID.

- Behavioral biometrics tools, which look at signals such as typing speed and patterns—as well as data from a device's accelerometer—combined with additional risk mitigation controls, such as device data and geolocation data, coupled with AI, can help detect synthetic identities and other anomalies quicker than other technologies.

User Experience Matters

Ensuring identity proofing is done properly is critical, but FIs don't want to make consumers jump through more hoops than necessary. FIs need to walk a fine line between user experience and security, which can be challenging. Conduct user feedback sessions to identify pain points in the identity verification process and refine systems accordingly.

- A financial institution that transitioned from active to passive liveness detection saw that the improved user experience led to an increase in application completion rates from 60% to over 95%, thereby reducing abandonment rates.[34]

- Banks have had to fine-tune models by balancing false rejection rates and false acceptance rates so that overly strict liveness detection models do not falsely reject legitimate users, causing frustration and customer support issues.

Regulatory and Privacy Considerations

States and different markets might have different legislation and regulatory requirements around different technologies. FIs need to conduct their own research to ensure compliance with any local policies or regulations. FIs should also monitor emerging legislation related to AI and biometrics to ensure compliance with future requirements.

- In the U.S., FIs need to be aware of state privacy laws that have specific requirements around biometrics and other data. Illinois's Biometric Information Protection Act (BIPA) requires consent and notification of how the data collector

[34] https://www.idrnd.ai/deep-dive-into-deepfakes/#:~:text=Deepfakes%20are%20rendered%20digital%20imagery,create%20deepfakes%20becomes%20increasingly%20accessible.

uses and stores the biometric information, in addition to other information. Colorado, Texas, and other states have implemented similar requirements.

- Outside the U.S., countries with strict data privacy laws (e.g., GDPR in Europe) require clear consent for biometric data processing.[35]

- U.S. banks need to adhere to authentication guidance from the Federal Financial Institutions Examination Council (FFIEC).[36] Of note, the FFIEC has not updated this guidance to account for the threat of deepfakes.

- Outside of the U.S., countries mandate alternative verification options for users who do not wish to use facial recognition.[37] While no such regulatory requirement exists in the U.S., 63% of individuals are wary of biometrics, and offering alternatives could be a useful approach.[38]

- The collection and storage of biometric data may raise significant privacy and data protection issues, requiring institutions to implement robust data security measures and comply with relevant regulations.[39]

Scalability and Deployment Challenges

- High demand for liveness checks can strain on-premises infrastructure or require manual verification from specialist fraud operations teams.

Policy

- Companies should consider policies governing the use of AI within an organization.

Maturity Model for Identity Controls to Combat Gen AI

Not all financial services organizations are created equal. Resource constraints in smaller institutions pose challenges that may be easier to solve for larger ones. In recognition of these differences, the following is a proposed Maturity Model for Identity Controls to Combat Gen AI.

The goal of this model is to lay out high-level technologies, ideas, and frameworks that FIs can pursue to mitigate Gen-AI-powered attacks. For example, on the technology side,

[35] https://commission.europa.eu/law/law-topic/data-protection_en

[36] https://www.ffiec.gov/sites/default/files/media/press-releases/2021/authentication-and-access-to-financial-institution-services-and-systems.pdf

[37] https://www.oaic.gov.au/privacy/australian-privacy-principles

[38] https://www.idtheftcenter.org/post/biometric-report-concerns-over-biometrics-use/

[39] https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57

starting with MFA, adding other risk-based technologies can help evolve protection against these attacks.

On the people side, it's a matter of making sure employees and consumers have the knowledge to spot potential attacks. Training, education, simulations, and testing are the best course of action to make sure employees and customers are prepared for potential attacks.

The standards and frameworks that FIs can use to help battle Gen AI are familiar ones. NIST SP 800-53 Rev. 5 is the standard for helping organizations across markets secure their systems. NIST SP 800-63-4 comprises the agency's Digital Identity Guidelines, which were recently updated to include information on liveness detection, mDLs, and phishing-resistant authentication. And the Cyber Risk Institute (CRI) has been working on specific guidance for FIs on AI.

This model can help FIs assess key technologies, people, standards, and frameworks to protect against these ever-present and ever-increasing threats.

Stage	Initial	Minimal	Evolving	Embedded
Technology	Multi-factor authentication (MFA), Data feed from ISAC or other service	Device Assessment, phishing-resistant MFA, Liveness detection, Data feed from ISAC or other service	Device Assessment, phishing-resistant MFA, Liveness detection, Data feed from ISAC or other service	Device Assessment, phishing-resistant MFA, Liveness detection, Data feed from ISAC or other service
People	Training, Acceptable Use Policy	Training, consumer, and employee education, Acceptable Use Policy	Training, consumer, and employee education simulation, Acceptable Use Policy	Training, simulation, testing, Acceptable Use Policy
Standards/ Framework	NIST SP800-53 Rev. 5, NIST SP800-63-4, NIST AI Risk Management Framework; CRI Profile, CRI FS AI RMF, FFIEC Guidance, CISA Cybersecurity Framework	NIST SP800-53 Rev. 5, NIST SP800-63-4, NIST AI Risk Management Framework; CRI Profile, CRI FS AI RMF, FFIEC Guidance, CISA Cybersecurity Framework	NIST SP800-53 Rev. 5, NIST SP800-63-4, NIST AI Risk Management Framework; CRI Profile, CRI FS AI RMF, FFIEC Guidance, CISA Cybersecurity Framework, OIDF Shared Signals Framework	NIST SP800-53 Rev. 5, NIST SP800-63-4, NIST AI Risk Management Framework; CRI Profile, CRI FS AI RMF, FFIEC Guidance, CISA Cybersecurity Framework, OIDF Shared Signals Framework

Conclusion

The threats of Gen AI to identity systems are significant. There will not be just one thing that FIs do to mitigate these threats. As attackers adapt, defenders will need to do the same. The ever-changing landscape will require constant monitoring and updates to keep pace with evolving threats.

FIs must prioritize investment in AI-driven defenses, employee training, and industry collaboration to stay ahead of these threats and be prepared to allocate the resources necessary to combat them and stay abreast so their systems can respond, as necessary. If they don't, fraud will climb, and customers may lose trust in the banking systems. Financial institutions must act now to implement robust, adaptive security measures to safeguard identity systems, protect consumers, and maintain trust in the financial ecosystem.