For release on delivery
11:45 a.m. EDT
April 17, 2025

Deepfakes and the AI Arms Race in Bank Cybersecurity

Remarks by

Michael S. Barr

Member

Board of Governors of the Federal Reserve System

at

Federal Reserve Bank of New York

New York, New York

April 17, 2025

Thank you for the opportunity to speak to you today about artificial intelligence (AI) and cybersecurity.[1] In the past, a skilled forger could pass a bad check by replicating a person's signature. Now, advances in AI can do much more damage by replicating a person's entire identity. This technology—known as deepfakes—has the potential to supercharge identity fraud. I've recently spoken about the importance of recognizing both the benefits and the risks of generative AI (Gen AI).[2] Today, I'd like to focus more on the darker side of the technology—specifically how Gen AI has the potential to enable deepfake technology, and what we should be doing now to defend against this risk in finance.

Escalating Threat of Gen-AI Facilitated Cybercrime

Cybercrime is on the rise, and cybercriminals are increasingly turning to Gen AI to facilitate their crimes. Criminal tactics are becoming more sophisticated and available to a broader range of criminals. Estimates of direct and indirect costs of cyber incidents range from 1 to 10 percent of global GDP.[3] Deepfake attacks have seen a twentyfold increase over the last three years.[4]

Cybercrime with deepfakes involves the same cat and mouse game common to sophisticated criminal activity. Both cybercriminals and financial institutions are

---

[1] The views expressed here are my own and are not necessarily those of my colleagues on the Federal Reserve Board or the Federal Open Market Committee.

[2] Michael S. Barr, "Artificial Intelligence: Hypothetical Scenarios for the Future" (speech at the Council on Foreign Relations, New York, NY, February 18, 2025), https://www.federalreserve.gov/newsevents/speech/barr20250218a.htm; Michael S. Barr, "AI, Fintechs, and Banks" (speech at the Federal Reserve Bank of San Francisco, San Francisco, CA, April 4, 2025), https://www.federalreserve.gov/newsevents/speech/barr20250404a.htm.

[3] International Monetary Fund, *Global Financial Stability Report*, chapter 3 (October 2024), See also, World Economic Forum, *Why We Need Global Rules to Crack Down on Cybercrime* (January 2023), https://www.weforum.org/stories/2023/01/global-rules-crack-down-cybercrime/.

[4] "Fraud attempts with deepfakes have increased by 2137% over the last three years," Signicat, February 20, 2025, https://www.signicat.com/press-releases/fraud-attempts-with-deepfakes-have-increased-by-2137-over-the-last-three-year#:~:text=Evolving20AI2Dbased20techniques20pose,AI2DDriven20Identity20Fraud20report.

constantly trying to outdo each other. Criminals develop new attack methods, and companies respond with better defenses. Here, the same technological innovations that enable the bad actors can also help those fighting cybercrime. However, there is an asymmetry—the fraudsters can cast a wide net of approaches and target a wide number of victims, and they only need a small number to be successful. Their marginal cost is generally low, and individual failures matter little. Conversely, companies must undergo a rigorous review and testing process to mount effective cyber defenses and will thus be slower in developing their defenses. A single failure is very costly. As we consider this issue from a policy perspective, we need to take steps to make attacks less likely by raising the cost of the attack to the cybercriminals and lowering the costs of defense to financial institutions and law enforcement.

Anatomy of a Deepfake

Deepfake attacks are those in which an attacker uses Gen AI to create a doppelganger with a person's voice or image and uses this doppelganger to interact with individuals or institutions to commit fraud. Deepfake technology is a particularly pernicious vehicle for cybercrime.[5] The process begins with voice synthesis, where Gen AI models can synthesize the speech of their victim not only in words, but also in phrase patterns, tone, and inflection. With just a short sample audio, for example, criminals assisted by Gen AI can impersonate a close relative in a crisis situation or a high-value bank client, seeking to complete a transaction at their bank.[6]

---

[5] Federal Bureau of Investigation, "Criminals Use Generative Artificial Intelligence to Facilitate Financial Fraud," public service announcement, December 3, 2024, https://www.ic3.gov/PSA/2024/PSA241203#:~:text=AI2DGenerated20Audio2C20aka20Vocal20Cloning&text=Criminals20generate20short20audio20clips,assistance20or20demanding20a20ransom.

[6] See note 5.

Criminals can also use Gen AI-generated videos to create believable depictions of individuals. For videos, Generative Adversarial Networks (GANs) are the core technology behind most deepfake systems.[7] GANs consist of two competing models, the generator and the discriminator, which compete with and improve each other. This competition results in increasingly realistic, indistinguishable fake images and videos.[8]

Deepfake technology can also be augmented by other AI tools; for instance, criminals can use AI to extract and organize extensive multimodal personal data to facilitate identity verification. Attackers can also turn to "dark web" tools, such as jailbroken versions of popular large language models, where the guardrails have been removed, to learn the deepfake trade and improve their attacks.[9]

Deepfakes in Action

I expect that many of you can recall examples of how deepfakes of politicians and prominent business executives have fooled the public and spread disinformation. Deepfakes are also being used to commit payment fraud. In one case in 2024, a sophisticated deepfake of the chief financial officer for British engineering and architectural firm Arup was reportedly deployed in a video meeting and convinced an Arup financial employee to transfer $25 million to thieves.[10]

---

[7] Tianxiang Shen, Ruixian Liu, Ju Bai, and Zheng Li, *"Deep Fakes" Using Generative Adversarial Networks (GAN)*, https://noiselab.ucsd.edu/ECE228_2018/Reports/Report16.pdf. McAfee, *Beware the Artificial Impostor* (May 2023), https://www.mcafee.com/content/dam/consumer/en-us/resources/cybersecurity/artificial-intelligence/rp-beware-the-artificial-impostor-report.pdf.

[8] "What is a GAN?" AWS, https://aws.amazon.com/what-is/gan/#:~:text=A20generative20adversarial20network20(GAN,from20a20database20of20songs.

[9] KELA, *The State of Cybercrime 2025 Report* (February 2025), https://www.kelacyber.com/resources/research/state-of-cybercrime-2025/.

[10] Kathleen Magramo, "British Engineering Giant Arup Revealed as $25 Million Deepfake Scam Victim," *CNN Business*, May 17, 2024, https://www.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk/index.html.

In another case, an attacker attempted to undertake a highly convincing audio deepfake of the chief executive of Ferrari, down to mimicking his southern Italian accent.[11] The recipient of the attack—another Ferrari executive—tested the caller with a personal question only the chief executive would know, which thankfully exposed the fraud.

And these institutions and individuals are not alone—a 2024 survey finds that over 10 percent of companies reported experiencing deepfake fraud attempts, and few steps have been taken to mitigate the risks.[12]

Particularly since COVID, we conduct much of our professional and personal lives over video. When we see realistic and interactive video images of a loved one in trouble, we are disposed to trust them and do what we can to help. Identity verification standards at banks often use voice detection, which may become vulnerable to Gen AI tools. If this technology becomes cheaper and more broadly available to criminals—and fraud detection technology does not keep pace—we are all vulnerable to a deepfake attack. These attacks can have significant financial costs to the victims of the crime and can also pose costs to society, eroding trust in communications and in institutions.

Defending Against Deepfakes

So what should we do? As I mentioned above, we should take steps to lessen the impact of attacks by making successful breaches less likely, while making each attack more resource-intensive for the attacker.

---

[11] Sandra Galletti and Massimo Pani, "How Ferrari Hit the Brakes on a Deepfake CEO," *MIT Sloan Management Review*, January 27, 2025, https://sloanreview.mit.edu/article/how-ferrari-hit-the-brakes-on-a-deepfake-ceo/.

[12] Chad Brooks, "1 in 10 Executives Say Their Companies Have Already Faced Deepfake Threats," business.com, June 28, 2024, https://www.business.com/articles/deepfake-threats-study/.

Let me start with ways to make successful breaches less likely. A key step is to recognize the importance of strong, resilient financial institutions in preventing attacks. Banks are frontline defenders against deepfake-enabled fraud due to their direct involvement with financial transactions and customer data. To verify payors, banks maintain identity verification processes, including multi-factor authentication and account monitoring practices. To the extent deepfakes increase, bank identity verification processes should evolve in kind to include AI-powered advances such as facial recognition, voice analysis, and behavioral biometrics to detect potential deepfakes. Other techniques focus on assessing the probability that AI has been used in audio or video based on underlying metadata and then flagging the identity or transaction for further review using other verification. These technical solutions can detect subtle inconsistencies in video and audio that human observers may miss.

Banks have two points of control over the transaction—confirming not only the sender's identity, but also the legitimacy of the recipient address. They can scrutinize the recipients of large or unusual transactions, employing advanced analytics to flag suspicious patterns that could indicate fraudulent activities, and perform additional reviews before authorizing a payment to a recipient that raises flags. Banks also invest in their human controls by maintaining up-to-date training for staff on the emerging risks and incorporating the necessary security measures to mitigate the damages from breaches when they occur. And they are engaging with other financial institutions to help define the threat and identify appropriate controls and mitigants.[13]

[13] See, for instance, FS-ISAC's report on deepfake threats and risk management at https://www.fsisac.com/hubfs/Knowledge/AI/DeepfakesInTheFinancialSector-UnderstandingTheThreatsManagingTheRisks.pdf.

Customers should do their part, enabling multi-factor authentication on their accounts and verifying unusual requests through a separate channel, even if the person making the request seems genuine. They should seek out education for themselves and their loved ones to help them detect and prevent fraud before it occurs.[14] And customers should value strong security practices at their financial institutions, including those which may add some friction to the user experience. The customers that may be the highest-value targets for criminals are often those with the largest digital presence, and thus most susceptible to deepfakes. They are also the customers who may prefer the most frictionless user experience, making detecting deepfakes more difficult. When it comes to protecting our money, we ought to expect and appreciate a little friction.

Regulators can help to reinforce the importance of cyber defenses in safe and sound banking through appropriate updates to guidance and regulation. As with all rules, we should be mindful of the impacts on smaller institutions and help ensure that rules are right-sized for the risk. In addition, we can work with core providers to understand the extent to which they are incorporating AI advancements in their products and services to help smaller banks defend against deepfakes and other emerging risks from the technology. Last, we can also highlight research and development for cybersecurity startups and research into tools to combat deepfakes and Gen AI-based fraud.

Regulators should consider how we could leverage AI technologies ourselves, including to enhance our ability to monitor and detect patterns of fraudulent activity at

---

[14] There are a variety of public and private resources that can help. See, for example, the National Security Agency/Central Security Service at https://www.nsa.gov/Press-Room/Press-Releases-Statements/Press-Release-View/Article/3523329/nsa-us-federal-agencies-advise-on-deepfake-threats/; and the National Cybersecurity Alliance at https://www.staysafeonline.org/articles/why-your-family-and-coworkers-need-a-safe-word-in-the-age-of-ai.

regulated institutions in real time. This could help provide early warnings to affected institutions and broader industry participants, as well as to protect our own systems.

In addition to preventing attacks, we should also explore ways of making attacks more costly. These may include coordination with domestic and global law enforcement, internationally consistent laws against cybercrime, and continued improvement on sharing threat intelligence and insights in real-time. The official sector and banks should continue efforts to improve fraud data sharing within the financial sector and help institutions respond more quickly to emerging Gen AI-driven threats. This will make it far harder for fraudsters to operate undetected, increasing the complexity and cost of their activities. But the sharing is only as good as the data, and banks must do their part. We should help ensure that banks and other regulated institutions meet their duties to report cyber incidents in a timely way, and regulators should too.[15]

Another way to disrupt the economics of cybercrime is by increasing penalties for attempting to use Gen AI to commit fraud and increasing investment in cybercrime enforcement. This includes targeting the upstream organizations that benefit from illegal action and strengthening anti-money-laundering laws to disrupt illicit fund flows and freeze assets related to cybercrime. The fear of severe legal consequences could help to deter bad actors from pursuing AI-driven fraud schemes in the first place.

Conclusion

Deepfakes are only one of many new techniques to facilitate cyberattacks, but they feel particularly salient because they are so personal. And they are on the rise.

---

[15] "Computer-Security Incident Notification Requirements for Banking Organizations and Their Bank Service Providers," 86 Fed. Reg. 66,424 (November 23, 2021), https://www.federalregister.gov/documents/2021/11/23/2021-25510/computer-security-incident-notification-requirements-for-banking-organizations-and-their-bank.

We will need financial institutions to adapt, collaborate, and innovate in the face of these emerging threats.

Thank you.