AFC

Fraud and Emerging Tech:

Deepfakes

April 2026

antifraudcollaboration.org

We welcome your feedback!

Please share your comments or questions at antifraudcollaboration.org/contact

As generative AI (genAI[1]) continues to advance, we have seen many beneficial and productive uses of this technology. However, alongside these gains, genAI can also be misused in harmful ways. For example, genAI has made it easier for bad actors to create sophisticated deepfakes, which can be used to defraud organizations. As such, it is important for organizations to consider the potential fraud risks arising from deepfakes and how to effectively mitigate such risks. This article, part of an emerging technology series from the Anti-Fraud Collaboration, explores what deepfakes are, how deepfakes may be used to perpetrate fraud against and within organizations, and how members of the financial reporting ecosystem can mitigate fraud risks arising from deepfakes.

1 For further discussion of genAI, refer to the Anti-Fraud Collaboration's report, Fraud and Emerging Tech: Generative AI.

What are deepfakes?

Deepfakes are a type of synthetic media created using AI that manipulates or generates content with the intent to deceive or demean others.[2] Deepfakes first emerged in the 2010s when bad actors used deep learning to create fake content (giving rise to the term "deepfakes").[3] Doing so required specialized knowledge and access to deep learning tools. Since then, the widespread availability of genAI has made it easier than ever for bad actors to create deepfakes. Also, as technology continues to improve, deepfakes are becoming increasingly sophisticated and difficult to identify. Examples of deepfakes include:

- **Text:** Text deepfakes mimic the writing style, tone, and messaging of an individual or organization to make false or fraudulent information more persuasive. Examples include text messages, emails, documents, and other written communications.

- **Image:** Image deepfakes are manipulated or fake images. Manipulations may include changing details of an existing photograph, such as swapping faces of individuals or altering backgrounds. Developments in AI image generation technology have also made it easy to generate realistic, but false, images from a simple text prompt. Image deepfakes often mimic an individual's likeness without consent or disclosure or depict fabricated scenes or events.

- **Video:** Like image deepfakes, video deepfakes are manipulated or fake videos. Video deepfakes can mimic an individual's facial expressions and body language and are often paired with audio deepfakes to create realistic but fake videos of someone speaking and acting. Video deepfakes can also portray events that have not occurred. Several AI video generation tools make it easy and accessible to generate realistic, but false, videos from a simple text prompt or a short sample of spoken video. Certain AI-enabled tools can also alter video in real time, enabling bad actors to appear as someone else during live interactions.

- **Audio:** Audio deepfakes mimic an individual's voice and can be used to impersonate someone during phone calls or voice messages. Certain AI-enabled tools allow real-time voice alteration, enabling bad actors to sound like someone else instantly.

It is important to note that the use of genAI tools to generate text, image, video, and audio is not inherently negative and that many constructive use cases exist. The key distinction between content created using genAI and deepfakes is that deepfakes **intend to deceive or demean**.

DEEPLIVE TECHNOLOGY

There is growing use of real-time deepfake technology (referred to as "deeplive" or "face swap" technology), which can be used to alter a person's appearance and voice instantly. Bad actors can use this technology to deceive individuals in phone or video calls and bypass identity verification technology.

2 Definition adapted from Deloitte's The Rise of Deepfakes: What Digital Platforms and Technology Organizations Should Know.
3 See the Department of Homeland Security's Increasing Threats of Deepfake Identities.

THE IMPACT OF GENAI ON DIGITAL DOCUMENTATION

At the 2025 Fraud Forum hosted by the Anti-Fraud Collaboration, stakeholders from across the financial reporting ecosystem discussed how advancements in technology, such as genAI and deepfakes, are reshaping the fraud risk landscape.

Participants emphasized that organizations now depend more on digitized manual signatures and digital documentation, such as invoices, contracts, and purchase orders, which are susceptible to generation and manipulation using genAI. As such, it is increasingly important for companies to think about the potential risk of document manipulation.

How can deepfakes be used to perpetrate fraud?

In addition to the broad reputational and societal harm that deepfakes can cause, companies must be aware that deepfakes can be used to commit fraud. Deepfakes can be used to perpetrate many forms of fraud, including:

- **Fraudulent financial transactions:** Video or audio deepfakes can be used to impersonate company leadership in video or phone calls to unsuspecting employees that appear to authorize financial transfers to third-party accounts, when in reality, these transfers are fraudulent.[4] Also, text deepfakes, such as fake invoices, contracts, or other agreements, can be used to appear to authorize or support inappropriate financial transactions.

- **Fraudulent claims evidence:** Image deepfakes can be used to falsify documentation, including photos, receipts, and other evidence to support the submission of fraudulent insurance claims, product warranty claims, or other forms of refund and reimbursement offered by companies that rely on photo evidence, thus resulting in the inappropriate distribution of funds.[5]

- **Synthetic identity fraud:** Synthetic identity fraud combines genuine personal information (such as a social security number) with additional pieces of fabricated data. Image or video deepfakes may be leveraged by bad actors to impersonate someone's likeness, supporting a synthetic identity. Synthetic identity fraud is particularly relevant to financial institutions, where bad actors use synthetic identities to bypass financial institutions' typical security and know-your-customer protocols.[6]

Fraudulent acts committed using deepfakes may be perpetrated by a third party with the intent to defraud a company and, therefore, may result in misappropriation of assets. However, there are also risks that those within a company could use deepfakes to facilitate misappropriation of assets or fraudulent financial reporting. Picture this:

- An employee uses image and text deepfakes to create fraudulent contracts, purchase orders, and invoices with a fake vendor to support the transfer of funds to an external bank account controlled by the employee.

- To meet earnings targets, management creates text and image deepfakes of contracts, purchase agreements, and shipping documents to support the recognition of revenue.

4 See Deloitte's Generative AI is expected to magnify the risk of deepfakes and other fraud in banking.
5 See KPMG's Deepfake Threats to Companies.
6 See Deloitte's Using biometrics to fight back against rising synthetic identity fraud.

Although these forms of fraud are not new, the use of deepfakes enables bad actors to perpetrate fraud with greater potency, at a much larger scale, and to target multiple victims using fewer resources.

DEEPFAKE ATTACKS IN THE NEWS

Deepfake attacks aren't just potential risks—they are happening right now:

- A finance employee at a multinational company received an email from someone whom he believed to be the CFO asking him to join a video call. The employee was originally suspicious of the email but joined the video call anyway. Several people were on the video call when he joined, all of whom looked and sounded like his colleagues. Based on the discussions on the call, the employee agreed to disburse the equivalent of $25,000,000 to a third-party bank account. Only later did the employee discover that the people who appeared on the video call were actually video deepfakes and that the $25,000,000 was inappropriately disbursed to a third party.
- The CEO of a subsidiary received a phone call from someone whom he believed to be the chief executive of the parent company. The caller requested that the subsidiary CEO transfer the equivalent of $243,000 to the bank account of a supplier. The subsidiary CEO grew skeptical of the call only after he transferred the funds and later discovered he had fallen victim to an audio deepfake so persuasive that it captured the parent company chief executive's accent and vocal patterns.

What are other risks related to deepfakes?

Bad actors can use deepfakes not only to perpetrate fraud for financial gain but also to damage an individual's or organization's reputation through the spread of misinformation.[7] For example, audio or video deepfakes could be used to impersonate a CFO discussing seemingly confidential, yet false, quarterly earnings before an earnings release to manipulate a company's share price. Image deepfakes could falsely depict inhumane working conditions or senior executives meeting with competitors with the objective of damaging the reputation of a company. Organizations should also be aware that bad actors are using deepfakes to conduct more sophisticated and targeted cyberattacks on organizations.[8] These examples may not result in direct financial loss to the company or financial statement misstatements, but they can lead to loss of stakeholder trust, loss of revenue, decreased share price, and other negative impacts to the organization.

THE LIAR'S DIVIDEND

The widespread awareness of deepfakes has given rise to a new risk: As the public becomes aware of deepfakes, they may trust authentic evidence less. Bad actors can claim that a piece of content is a deepfake to avoid accountability and reputational harm, when in reality, the content is credible. This phenomenon is called the liar's dividend.[9]

The prevalence of deepfakes and the concept of the liar's dividend call for increased attention to content. Although the primary risk of misinformation lies with deepfakes themselves, it is important for organizations to be cognizant of the liar's dividend. If an organization faces an accusation of improper conduct, presenting real evidence to refute it might not be enough to convince the public that the claims are false, given the widespread availability of technology to create misleading content.[10]

7 See further discussion at FS-ISAC's Deepfakes in the Financial Sector: Understanding the Threats, Managing the Risks.
8 See KPMG's Deepfake Threats to Companies.
9 The term was coined by Bobby Chesney and Danielle Citron in their study *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*.
10 See further discussion at The Liar's Dividend & What Corporate Leaders Can Learn From GenAI's Impact on Election Day.

How can stakeholders throughout the financial reporting ecosystem mitigate risks arising from deepfakes?

According to a 2025 Gartner survey, 62% of organizations experienced a deepfake attack in the past 12 months. As such, it is critical that organizations consider strategies to mitigate the risks deepfake attacks present. Deepfake attacks can target any member of an organization, from junior staff to seasoned executives and board members. Accordingly, effective mitigation is a shared responsibility, and stakeholders throughout the financial reporting ecosystem have a role to play:

BOARDS OF DIRECTORS

- **Build the organization's trust reserve.**[11] In their oversight role, boards can work with management to continue to build trust within the organization and externally. This includes evaluating trust metrics, such as media tone, stakeholder and investor confidence, and internal morale. Increased trust can help mitigate the risk of reputational damage arising from a potential deepfake attack.

- **Plan and rehearse responses to potential attacks.**[12] A best practice for boards is to rehearse crisis management protocols with management to ensure they are aware of the procedures to be followed in the event of a deepfake attack, especially attacks targeted at senior executives.[13] Such protocols may include, among other things, engaging counsel or a third-party specialist to investigate incidents. Rehearsing protocols through scenario analyses and tabletop exercises can help decrease response time in the event of a deepfake attack and help minimize impact on the organization.

MANAGEMENT

- **Increase awareness across the organization.**[14] Management has an important role in setting the tone at the top of the organization, encouraging critical thinking, and acting with integrity. Also, because deepfakes can affect any member of the organization, it is critical to provide training to all employees to keep them informed of the risks of deepfakes, common indicators of deepfakes to be aware of, and procedures to follow if an employee believes they have encountered or fallen victim to a deepfake attack.

- **Define "normal" behavior.**[15] Several deepfake attacks resulting in fraudulent financial transactions have been successful because employees believed and acted on a request made to them that was outside of standard processes and controls. For example, to

A best practice for boards is to rehearse crisis management protocols with management to ensure they are aware of the procedures to be followed in the event of a deepfake attack, especially attacks targeted at senior executives.

11 See further discussion at the National Association of Corporate Directors' The Founding Fathers Used Their Real Names, But Deepfakes Use Yours.
12 See further discussion at the Department of Defense's Contextualizing Deepfake Threats to Organizations.
13 See further discussion at The Founding Fathers Used Their Real Names, But Deepfakes Use Yours.
14 See further discussion at A Fraud Forum Hosted by the AFC: Navigating a Dynamic Fraud Risk Environment.
15 Ibid.

address the risk that employees may believe a fraudulent request from management to transfer money, management can send a clear message to employees that standard processes will always be followed if any fund transfers are requested and that there are certain requests that management will never make to an employee. Management can also create an environment where employees are encouraged to directly verify nonstandard requests through a separate communication channel.[16]

INTERNAL AUDITORS

- **Conduct risk assessments to identify deepfake vulnerabilities.**[17] Internal audit can conduct risk assessments to proactively identify potential deepfake vulnerabilities within the organization. These risk assessments can then be used to help the organization close any identified gaps and refine the incident response plan.

- **Strengthen internal controls.**[18] Internal auditors can collaborate with business process owners to strengthen internal controls, specifically surrounding processes involving external vendors and customers. For example, internal auditors may consider the need to advise process owners to introduce multifactor authentication to processes involving remittance of funds to external vendors to mitigate the risk of one individual falling victim to a deepfake scheme. Incorporating biometric authentication to mitigate the risk of synthetic identity fraud may also be considered, where appropriate.

EXTERNAL AUDITORS

- **Consider the potential impact of deepfake attacks on fraud risk assessment.** External auditors consider two types of material misstatements relevant to fraud—those arising from fraudulent financial reporting and those arising from misappropriation of assets. Depending on the nature of the company's business, external auditors may consider how deepfake attacks perpetrated by a third party could potentially result in a material misstatement of the financial statements arising from misappropriation of assets. Auditors may also consider how management or others within the entity could use deepfakes to perpetrate or conceal fraudulent financial reporting or misappropriation of assets, resulting in a material misstatement of the financial statements. Three conditions are generally present when fraud occurs—incentive, opportunity, and rationalization. Deepfakes may create more opportunities for fraud when an incentive and ability to rationalize committing a fraudulent act already exist.

- **Involve forensic specialists when needed.** If a deepfake attack has occurred, the company may engage counsel or a third-party specialist to understand the nature and impact of the attack. In such cases, the auditor may consider the need to involve a forensic specialist, who can aid the auditor in assessing whether the third-party investigation was sufficient and evaluating the related findings to determine whether there is a material misstatement of the financial statements and any other impacts on the audit.

Deepfakes may create more opportunities for fraud when an incentive and ability to rationalize committing a fraudulent act already exist.

16 See further discussion at EY's When Deepfake Dangers Cause Real Crises.
17 See further discussion at the Institute of Internal Auditors' All Things Internal Audit: Deepfakes and AI Fraud Risks for Internal Auditors.
18 See further discussion at the Institute of Internal Auditors' The Fraudsters Have AI, Too.

No matter an individual's role within the financial reporting ecosystem, it is important to remain vigilant for potential deepfakes when evaluating content. Although deepfake detection technologies are emerging, human recognition remains critical in mitigating the risks. Recent evaluations of deepfake detection technology indicate that, although many different tools are available, the effectiveness of such tools in identifying deepfakes is mixed.[19] To promote increased skepticism, organizations can consider implementing the SIFT framework to help determine if content is credible and reliable.[20] The SIFT framework includes the following steps:

1. **Stop.** Before acting on content received, stop. Ask yourself whether you recognize the source of the information and its reputation.

2. **Investigate the source.** Is the source from which you've received the content reliable? Consider the email address or phone number from which any text, videos, audio, or images are received.

3. **Find better coverage.** Validate the messaging. Did you receive an email with a video from the CEO sharing information that is inconsistent with normal messaging from leadership? Or maybe you've received a spontaneous phone call from the CFO asking you to disburse funds to a new bank account? Find other sources or use other communication channels to validate the content received before taking action.

4. **Trace.** Understanding the context is key. When evaluating content, trace the messaging of the content back to its original source, where possible, to better understand the original context. Verified sources also risk falling victim to deepfakes; therefore, tracing the content to its original source to obtain the full context can help determine whether content is a deepfake.

In this dynamic environment, it is critical for organizations to keep current on advancements in technology and continuously adapt fraud mitigation strategies.

Conclusion

Deepfakes are just one example of how technology is changing the fraud risk landscape by increasing the speed, scale, and sophistication of fraud schemes. In this dynamic environment, it is critical for organizations to keep current on advancements in technology and continuously adapt fraud mitigation strategies. The AFC will continue to monitor emerging technologies affecting the fraud risk environment and will work with stakeholders across the financial reporting ecosystem to advance awareness, dialogue, and mitigation efforts.

19 For example, in *Deepfakes: Real Threat*, KPMG noted that more than 100,000 models have been developed to create and/or detect deepfakes; however, only around 3,000 of these models can detect deepfakes properly, with most others able to detect only unsophisticated deepfakes.
20 The SIFT framework was developed by digital literacy expert, Mike Caulfield. See further information here.