

# AI Deepfakes

When Seeing and Hearing Can't be Trusted



# Key Takeaways

## More than just a menace

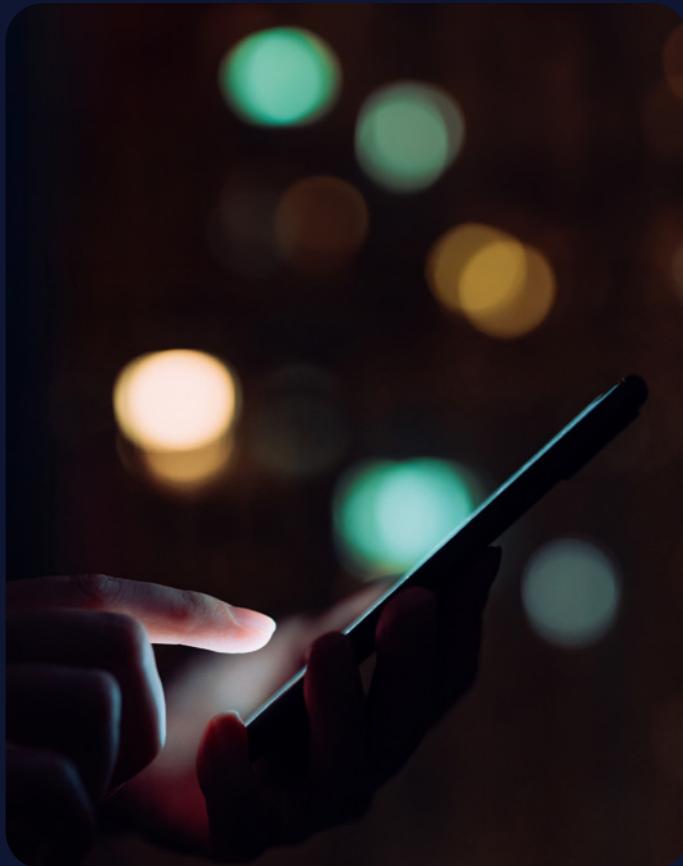
Deepfakes have evolved from social media entertainment to powerful tools of deception, with nearly 5% of all fraud attempts in 2024 estimated to involve deepfake elements.

## Workplace infiltration

AI and deepfakes are also infiltrating the workplace. At Money20/20 USA (October 2025), a leading technology firm revealed that up to 50% of job applications it receives are fake.

## Trust no-one

Traditional identity checks are no longer sufficient and the focus must shift to zero-trust communication and training for employees and law enforcement to stay ahead of sophisticated fraud techniques.



*1 in 4*

Estimated number of fake job candidate profiles by 2028<sup>1</sup>

*320*

Estimated number of companies infiltrated by North Korean IT workers in the past year<sup>2</sup>

*\$10.5 tn*

Estimated global annual cost of cybercrime in 2025, up from \$3 trillion in 2015<sup>3</sup>

## AI Deepfakes

Deepfakes are the new face of deception. They are voices, images, videos or even text created by artificial intelligence (AI) that look and sound indistinguishably real. Once mere entertainment novelties, they have evolved into powerful tools of manipulation and fraud, marking a new era in financial crime.

This wave of AI-driven deception is now infiltrating the workplace and recruitment processes. Some estimates suggest that by 2028, one in four candidate profiles worldwide could be fake.<sup>4</sup> Meanwhile, one technology company the Citi Institute spoke with at Money 20/20 USA (October 2025) told us that 50% of job applications it receives are fake.

Identity deception is accelerating. AI improvements enable deepfakes to mimic real people and make it difficult to detect synthetic ones. The rise of video-based hiring and remote work, especially in the technology and web3 sectors, further amplifies the risk.

As recruitment interviews shift to virtual environments, imposters with fake credentials may be able to secure jobs. The danger lies not just in what they fake, but in how long they can stay undetected, infiltrating sensitive systems and installing malware or ransomware.

## State-sponsored Deepfake Employees

A striking example is the surge in state-sponsored actors using deepfake technology to infiltrate global companies. These candidates are either completely AI-generated or have their appearance significantly altered using deepfake technology.

North Korea has emerged as a hotspot, with its operatives often posing as IT professionals aiming to infiltrate foreign companies using false identities. An estimated 320 companies have been infiltrated by North Korean IT workers in the past year.<sup>5,6</sup>

The deepfake employee scam tends to be a long-term campaign involving multiple individuals across locations including overseas operators (impersonators), onshore collaborators (mules), brokers, and money movement handlers. A typical operation includes:

- Targeting remote roles with limited in-person onboarding
- Creating synthetic identities with fabricated CVs, social profiles, and realistic headshots or videos
- Participating in deepfake interviews using voice or video cloning, providing fabricated work history and personal references

- Establishing an onshore foothold through a rented address or a local collaborator for administrative purposes and deliveries
- Having company laptops or authenticators mailed to the onshore address
- Connecting remotely to company systems using the provided credentials and hardware, often with the help from the collaborator
- Co-ordinating multi-location teams to scale the operation

Once candidates gain access to the company ecosystem, the scam goes beyond simple deception. Potential impacts include the infiltration of sensitive systems and leakage of trade secrets, theft of intellectual property and customer data, data breaches and ransomware attacks, generation of foreign currency revenues for the sponsoring regime, and the erosion of trust and reputation.

“

For one job posting alone, we received over 800 applications in a matter of days. When we conducted a deeper analysis of 300 candidate profiles, over one-third were outright fraudulent. These weren't just candidates exaggerating their experience – these were entirely fabricated identities, many leveraging AI-generated resumes, manipulated credentials, and, most concerning, deepfake video interviews.

Vijay Balasubramaniyan, Co-Founder and CEO, Pindrop Security

”

This is not just a problem for large corporates. Small and mid-sized businesses, which often lack the resources to detect sophisticated hiring fraud, are particularly vulnerable.

Corporates often fail to conduct thorough verification of remote hires and contractors. Standard background checks often rely on self-reported information or basic identity verification, which can be manipulated through deepfake visuals, fabricated credentials, or stolen digital identities.

Deepfakes are not confined to job applicants. They also impact senior leadership, customers, and suppliers. In the financial sector, this could extend to synthetic identity creation, fraudulent transaction authorization, and automated money transfer scams.

### Financial Deepfake Fraud

There have been several single-event, high-impact financial frauds where deepfakes impersonated trusted executives to authorize or redirect payments. Unlike systemic infiltration cases, the objective here is immediate financial gain rather than long-term access.

In a widely reported case in 2024, a UK multinational became the target of a strikingly sophisticated fraud. An employee of its Hong Kong office received what appeared to be a video call from the chief financial officer and other senior leaders, urging urgent money transfers for a confidential transaction.<sup>7</sup>

In reality, the voices and images of the executives had been deepfaked and every participant on the call except the employee was a synthetic representation. The employee proceeded to make 15 separate transfers totalling around HKD200 million (\$25 million) to an offshore account before the deception was uncovered.

Several such incidences have been reported in recent years, with potentially more going unreported. Such incidents illustrate the danger of deepfakes, rendering traditional verification controls such as voice recognition and visual confirmation inadequate.

It also highlights the importance of continuous education to understand the common techniques that fraudsters use to enable the early detection of fakes. More importantly, ongoing and proactive education is key for law enforcement to stay ahead of the perpetrators.

### Audio Spoofs to Full-Motion Real Time Video

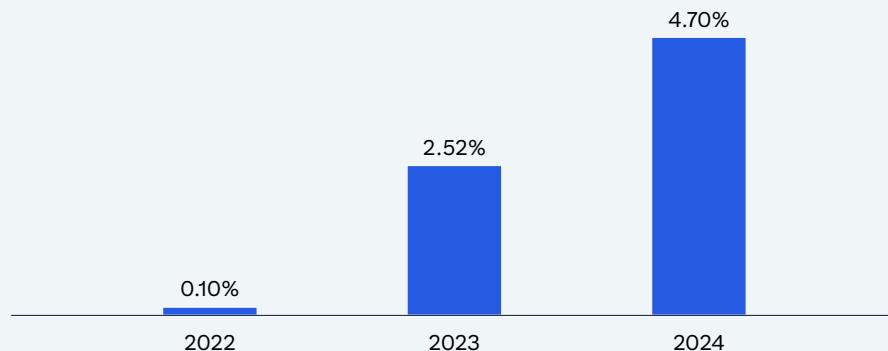
Deepfake fraud attempts are multiplying fast. 2024 saw a marked increase in deepfake-related fraud, accounting for nearly 4.7% of all fraud attempts, up from 0.10% in 2022.<sup>8</sup>

Deepfake fraud attempts vary by sector: in consumer credit they represent nearly 12% of all fraud, followed by real estate (about 8%) and payments (5%). By contrast, fewer deepfake attempts occur where the perceived gains are lower, such as education.<sup>9</sup>

Up to 8 million deepfakes are expected to be shared online by the end of 2025, up from 500,000 in 2023, suggesting a doubling every six months.<sup>10</sup> Easy access to powerful AI tools and vast amounts of data are likely contributing factors.

A survey of over 100 fraud executives from global financial institutions over 2Q2025-3Q2025 suggest greater anticipation of increased fraud losses in banking payments over the next three years, especially for the U.S.<sup>11</sup> The increase is attributed to growth in AI-powered deepfake attacks and synthetic identity fraud.

Figure 1. Deepfake fraud as percentage of all fraud attempts



Source: Signicat VideoID data, full-year 2021-2024, from "The Battle in the Dark"

**Figure 2.** Key forces behind the growing speed and sophistication of speech-based deepfakes<sup>12</sup>



### Use of automated bots

Previously, speech generation tools had a 4-7 second delay between input and synthetic voice output. Today, LLMs have reduced that delay to near real-time. This makes it increasingly difficult to distinguish synthetic voices from real ones.



### Emotional-sounding AI

Advances in synthetic speech have enabled text-to-speech voices to convey emotions like joy, anger, empathy, and sadness. AI models can now learn and imitate emotional tones from human speech, making these synthetic voices even more convincing.



### Real-time voice conversion

Companies have created tools for real-time voice conversion, allowing users to change pitch, timbre, and accent instantly. While this technology benefits voice dubbing, gaming, and content creation, it also makes it easier for fraudsters to evade voice recognition systems by masking their voice.

*Source: Pindrop 2025 Voice Intelligence and Security Report, Citi Institute*

**Voice deepfakes:** Voice deepfakes are synthetic audio created to sound like the target individual. They are created by training a model on samples of a person's voice and then providing a text to generate the fake speech.

Combining large language models (LLMs) with text-to-speech engines can enable voice bots to respond in real time. The LLM generates real-time answers to questions, and the text-to-speech engine vocalizes them, even conveying emotions such as empathy or urgency.

**Video deepfakes:** These extend the threat to full visual manipulation. Initially, video deepfakes were pre-recorded content that was manipulated before distribution.

Now advanced deepfakes can manipulate someone's likeness using generative adversarial networks and other machine learning techniques designed to create lifelike synthetic media.

Attackers often use publicly available data like photos, videos, and audio from social media websites or corporate filings to build the model. During a live call, the

model overlays the synthetic likeness in real time, replicating facial cues, voice, and gestures.

**Hybrid deepfakes:** Many live deepfake scams use a hybrid strategy, combining multiple types of deepfakes (including voice and video), with traditional social engineering tactics, fabricated documents, and credential theft.

Instead of a one-off scam, such deepfakes are often associated with long-term planned infiltration campaigns. They are designed to build trust, before compromising systems or extracting sensitive data.

### Detecting Deepfakes is Getting Harder

Detecting deepfakes is getting harder, especially audio. Early versions often contained noticeable pauses as the operator typed responses. Recent iterations eliminate these flaws, producing seamless and natural-sounding speech.

The number of generative AI (GenAI) systems capable of cloning voice and video has surged, rising from roughly 100-150 tracked systems last year to more than 500 today.<sup>13</sup> Much of this growth is driven by open-source tools that are easier to access, cheaper to run, and require less data to create highly convincing fakes.

There are cases of bots simulating empathy. For example, a bot may remark during an interaction "It must have been a long day. So how are you holding up?" This carefully engineered scripted empathy, helps increase credibility and makes the interaction appear more convincing.

In contrast, video deepfakes still show some tell-tale flaws such as blurriness or unnatural pixelation. But the technology is evolving fast.

## How are Corporates Responding?

The global annual cost of cybercrime is estimated to reach \$10.5 trillion by 2025, up from \$3 trillion in 2015.<sup>14</sup> Old school fraud using one-time passwords (OTPs) and phishing continue to exist, but deepfakes are seeing a rapid increase.

Corporates are responding in different ways. Many firms are prioritizing the detection of C-suite impersonation due to the high-profile nature of these attacks targeting CEOs and CFOs. Others are investing in tools to safeguard video communications more broadly.

As fraud grows more sophisticated, traditional identity checks such as document scans or liveness tests are no longer sufficient. Trust cannot be established through a single interaction.

Verification must evolve into multi-layered digital constructs that combine biometrics, behavior and device data, and contextual cues. The notion of continuous identity is becoming crucial.

## Zero-Trust Communication is Essential

In a GenAI-powered world, trust can no longer be assumed, it must be continuously verified.

Every interaction, whether from inside or outside the network, must be verified

through multiple layers of identity, device, and behavioural validation. The principle must be “Never trust, always verify”.

While corporates are redesigning call centres for zero trust, communication channels like phone calls and emails remain dangerously outdated. Many corporate systems still rely on voice recognition, caller ID, and email domain as proof of authenticity.

The next frontier is zero-trust communication, where every conversation and message undergo real-time authentication using biometric voiceprints, behavioral analytics, and device-level identity tokens. Likewise, email security must move towards cryptographic message signing, AI-based anomaly detection, and intent verification.

## Fighting AI with AI

The way financial services combat fraud will fundamentally change as criminals adopt AI to perpetrate scams. Deepfakes' ability to circumvent traditional defenses illustrate this shift. While the pace of AI-driven fraud is alarming, the fight against deepfakes is winnable.

The same AI technology that enables fraud, can also be used against it. Advanced AI agents are now capable of mapping scam networks, flagging

manipulated audio and video, and intercepting social engineering attempts with increasing precision. Building AI-driven defense systems is becoming as critical to financial security as cybersecurity firewalls.

As AI agents evolve and operate autonomously, the risks escalate. Bad actors can deploy agents at scale to impersonate senior executives, manipulate employees, or mislead customers. This raises the bar for verification. Financial institutions must move beyond validating users to also verifying the identity, intent, and provenance of AI agents.

In *Citi GPS: Agentic AI* we highlight several examples of how AI is being used to counter fraud. One leading global bank, for instance, integrated real-time deepfake detection into its call center infrastructure. The detection process happens seamlessly without introducing latency or disrupting the natural flow of conversation. The AI tool analyses the audio stream and flags signs of synthetic manipulation.

The industry is also beginning to build frameworks such as Know Your Agent (KYA), mirroring the KYC standard, to safeguard trust in digital interactions.



## Authors



**Ronit Ghose**  
Global Head, Future of Finance, Citi Institute  
ronit.ghose@citi.com



**Sophia Bantanidis**  
Future of Finance,  
Citi Institute  
sophia.bantanidis@citi.com



**Prag Sharma**  
Future of Finance,  
Citi Institute  
prag.sharma@citi.com



**Kaiwan Master**  
Future of Finance,  
Citi Institute  
kaiwan.hoshang.master@citi.com



**Ronak Shah**  
Future of Finance,  
Citi Institute  
ronak.sharad.shah@citi.com

## Contributors

### Vijay Balasubramaniyan

Pindrop Security

## Endnotes

<sup>1</sup> Gartner, Gartner Survey Shows Just 26% of Job Applicants Trust AI Will Fairly Evaluate Them, 31 July 2025.

<sup>2</sup> Fortune, North Korean IT Worker Infiltrations Exploded 220% Over the Past 12 months, with GenAI Weaponized at Every Stage of the Hiring Process, 04 August 2025; CrowdStrike, Threat Hunting Report, 2025.

<sup>3</sup> eSentire, Cybercrime to Cost the World \$9.5 Trillion USD Annually in 2024.

<sup>4</sup> Gartner, Gartner Survey Shows Just 26% of Job Applicants Trust AI Will Fairly Evaluate Them, 31 July 2025.

<sup>5</sup> Fortune, North Korean IT Worker Infiltrations Exploded 220% Over the Past 12 months, with GenAI Weaponized at Every Stage of the Hiring Process, 04 August 2025.

<sup>6</sup> CrowdStrike, Threat Hunting Report, 2025.

<sup>7</sup> CNN Business, British Engineering Giant Arup Revealed as \$25 million Deepfake Scam Victim, 17 May 2024.

<sup>8</sup> Signicat, The Battle in the Dark, October 2025.

<sup>9</sup> Signicat, The Battle in the Dark, October 2025.

<sup>10</sup> UK Government (UK.Gov), Innovating to Detect Deepfakes and Protect the Public, 05 February 2025.

<sup>11</sup> Datos Insights, Five Forces Disrupting Global Fraud Prevention by 2030, October 2025.

<sup>12</sup> Pindrop 2025 Voice Intelligence and Security Report.

<sup>13</sup> Citi Institute Future of Finance Forum 2025 Video, Deep Dive into Deepfakes, 09 July 2025.

<sup>14</sup> eSentire, Cybercrime to Cost the World \$9.5 Trillion USD Annually in 2024.

For digital accessibility support, please contact your Citi Team for immediate queries, or email [accessibility@citi.com](mailto:accessibility@citi.com). You may visit [Accessibility at Citi | Citi.com](#) for more information

## IMPORTANT DISCLOSURES

This communication has been prepared by Citigroup Global Markets and is distributed by or through its locally authorised affiliates (collectively, the "Firm"). This communication is not intended to constitute "research" as that term is defined by applicable regulations, though it may contain thematic content that has been or may be contained in research reports. Unless otherwise indicated, any reference to a research report or research recommendation is not intended to represent the whole report and is not in itself considered a recommendation or research report. The views expressed by each author herein are their personal views and do not necessarily reflect the views of their employer or any affiliated entity or the other authors, may differ from the views of other personnel at such entities, and may change without notice. You should assume the following: The Firm may be the issuer of, or may trade as principal in, the financial instruments referred to in this communication or other related financial instruments. The author of this communication may have discussed the information contained herein with others within the Firm and the author and such other Firm personnel may have already acted on the basis of this information (including by trading for the Firm's proprietary accounts or communicating the information contained herein to other customers of the Firm). The Firm performs or seeks to perform investment banking and other services for the issuer of any such financial instruments. The Firm, the Firm's personnel (including those with whom the author may have consulted in the preparation of this communication), and other customers of the Firm may be long or short the financial instruments referred to herein, may have acquired such positions at prices and market conditions that are no longer available, and may have interests different or adverse to your interests. This communication is provided for information and discussion purposes only. It does not constitute an offer or solicitation to purchase or sell any financial instruments. The information contained in this communication is based on generally available information and, although obtained from sources believed by the Firm to be reliable, its accuracy and completeness is not guaranteed. Certain personnel or business areas of the Firm may have access to or have acquired material non-public information that may have an impact (positive or negative) on the information contained herein, but that is not available to or known by the author of this communication. The Firm shall have no liability to the user or to third parties, for the quality, accuracy, timeliness, continued availability or completeness of the data nor for any special, direct, indirect, incidental or consequential loss or damage which may be sustained because of the use of the information in this communication or otherwise arising in connection with this communication, provided that this exclusion of liability shall not exclude or limit any liability under any law or regulation applicable to the Firm that may not be excluded or restricted. The provision of information is not based on your individual circumstances and should not be relied upon as an assessment of suitability for you of a particular product or transaction. Even if we possess information as to your objectives in relation to any transaction, series of transactions or trading strategy, this will not be deemed sufficient for any assessment of suitability for you of any transaction, series of transactions or trading strategy. The Firm is not acting as your advisor, fiduciary or agent and is not managing your account. The information herein does not constitute investment advice and the Firm makes no recommendation as to the suitability of any of the products or transactions mentioned. Any trading or investment decisions you take are in reliance on your own analysis and judgment and/or that of your advisors and not in reliance on us. Therefore, prior to entering into any transaction, you should determine, without reliance on the Firm, the economic risks or merits, as well as the legal, tax and accounting characteristics and consequences of the transaction and that you are able to assume these risks. Financial instruments denominated in a foreign currency are subject to exchange rate fluctuations, which may have an adverse effect on the price or value of an investment in such products. Investments in financial instruments carry significant risk, including the possible loss of the principal amount invested. Investors should obtain advice from their own tax, financial, legal and other advisors, and only make investment decisions on the basis of the investor's own objectives, experience and resources. This communication is not intended to forecast or predict future events. Past performance is not a guarantee or indication of future results. Any prices provided herein (other than those that are identified as being historical) are indicative only and do not represent firm quotes as to either price or size. You should contact your local representative directly if you are interested in buying or selling any financial instrument, or pursuing any trading strategy, mentioned herein. No liability is accepted by the Firm for any loss (whether direct, indirect or consequential) that may arise from any use of the information contained herein or derived herefrom. Although the Firm is affiliated with Citibank, N. A. (together with its subsidiaries and branches worldwide, "Citibank"), you should be aware that none of the other financial instruments mentioned in this communication (unless expressly stated otherwise) are (i) insured by the Federal Deposit Insurance Corporation or any other governmental authority, or (ii) deposits or other obligations of, or guaranteed by, Citibank or any other insured depository institution. This communication contains data compilations, writings and information that are proprietary to the Firm and protected under copyright and other intellectual property laws, and may not be redistributed or otherwise transmitted by you to any other person for any purpose. © 2025 Citigroup Global Markets Inc. Member SIPC. All rights reserved. Citi and Citi and Arc Design are trademarks and service marks of Citigroup Inc. or its affiliates and are used and registered throughout the world.