VIEW POINT

THE IMPACT OF AI-GENERATED DEEPFAKES ON FINANCIAL FRAUD

Abstract

AI-generated deepfakes are becoming a significant challenge in financial fraud prevention, enabling sophisticated scams that manipulate biometric authentication, voice recognition, and video verification systems. As deepfake technology evolves, businesses and financial institutions must adopt robust deepfake detection strategies to mitigate risks. Explore the impact of deepfakes on financial fraud, the latest detection methods, and proactive measures organisations can take to safeguard their operations against fraudulent AI-generated content.

Infosys®
Navigate your next

A fraudster used an audio deepfake to convincingly mimic a CFO in a virtual meeting, successfully bypassing security checks and exploiting organisational trust. The result? A $25 million unauthorised transfer. This incident highlights the growing threat of deepfake-driven fraud, where AI-powered deception targets executives with financial authority and sensitive access.

Deepfake technology is advancing rapidly, and the cost of producing high-quality deepfake content is decreasing. This shift is expected to lead to a more complex threat landscape, where attacks will become more frequent and successful.

To effectively counter and mitigate these threats, it will be crucial for governments, industries, and societies to collaborate and coordinate their efforts.

This article explores the impact of AI-generated deepfakes on financial fraud, the challenges of deepfake detection, and the strategies organisations must adopt to safeguard their assets.

The rise of deepfake technology

Deepfake technology, powered by artificial intelligence, leverages sophisticated algorithms like Generative Adversarial Networks (GANs) to create hyper-realistic audio, video, or images that mimic real people. Originally a tool for entertainment, its accessibility has skyrocketed. Open-source software and cloud computing have democratised deepfake creation, enabling even amateurs to produce convincing fakes with minimal resources. Unlike traditional machine learning, where models improve with structured input, deep learning algorithms autonomously identify and refine features.

Deepfakes present a growing threat to enterprises, society, and data privacy. The World Economic Forum highlights them as a top risk due to their potential to undermine trust in media and institutions worldwide.

For financial fraudsters, this technology is a goldmine. While traditional scams rely on phishing emails or forged signatures, AI-generated deepfakes offer a visceral, real-time avenue to deceive.

According to Sumsub, with a 245% year-on-year increase in detections from Q1 2023 to Q1 2024, deepfake incidents are rapidly gaining global attention. The top three countries experiencing the highest year-on-year growth in deepfake detections are China (2,800%), South Korea (1,625%), and Singapore (1,100%).

The rise of AI-driven fraud is staggering:

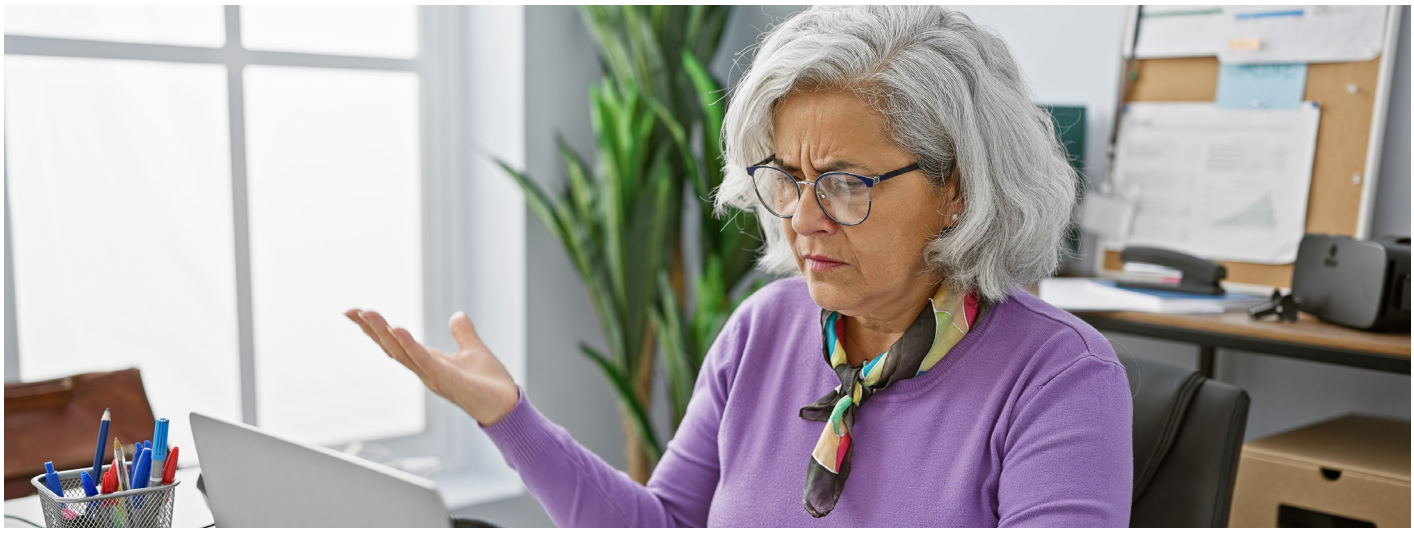
- 40% of all deepfake attacks target the financial sector, making it one of the most affected industries.
- Nearly half of all businesses have experienced fraud involving audio or video deepfakes, with average losses per incident nearing $450,000.

How deepfakes work: The mechanics of deception

There are enough digital traces available online for bad actors to easily create realistic deepfakes. These models generate synthetic media by mapping facial movements, voice patterns, and even emotional cues onto a target. The manipulated visuals are then streamed through virtual cameras, replacing real webcam feeds and making it nearly impossible to distinguish between real and fake participants. The result is a forgery so lifelike that even trained professionals struggle to spot the difference. Fraudsters pair this with social engineering. They might scrape LinkedIn for executive profiles, harvest voice samples from earnings calls, or use publicly available photos to craft their deepfake. For instance, a CFO's voice can be cloned from a 30-second podcast clip and then used to authorise a multi-million-dollar transfer. The speed and precision of these attacks leave little room for hesitation or verification, exploiting the fast-paced nature of financial decision-making.

The impact: Understanding the risks and stakes

Deepfake-driven fraud can compromise brand integrity, damage reputations, and spread misinformation, leading to financial implications and legal non-compliance. Here are the types and key assets at risk, common impersonation targets, and primary recipients of deepfake attacks within the financial sector.



Types of risks posed by deepfakes

Deepfake technology introduces multifaceted risks to financial institutions, including:

Fraudulent Transactions

Criminals use synthetic media (video, audio, or text) to impersonate executives, clients, or trusted third parties, tricking employees into approving unauthorised transfers (e.g., Business email compromise scams).

Reputational damage

Fake videos/audio of executives making inflammatory remarks or false announcements can tank stock prices or erode customer trust.

Operational disruption

Deepfakes could bypass voice or facial recognition authentication systems, compromising account security.

Regulatory and compliance penalties

Institutions failing to detect deepfake-driven breaches may face fines for inadequate financial fraud prevention measures.

Market manipulation

Fabricated news clips or executive statements could artificially inflate/deflate asset values.

Assets at risk

Financial institutions' most vulnerable assets include:

Financial reserves

Direct theft via fraudulent transfers (e.g., the $25M case described above)

Customer data

Deepfakes enable social engineering to extract sensitive information (e.g., account credentials, national identification numbers)

Intellectual property

Fake internal communications could coerce employees into leaking proprietary algorithms or strategies

IT infrastructure

Synthetic media might trick IT teams into granting system access or disabling security protocols

Brand equity

Loss of customer trust due to perceived security failures or reputational harm

Impersonation targets and methods

Deepfake attackers typically impersonate:

C-suite executives

(e.g., CEOs, CFOs): Miscreants mimic high-level authorities to approve large transactions or policy changes.

IT/compliance staff

Fraudsters pose as internal tech teams to phish credentials or bypass multi-factor authentication.

Clients or vendors

Spoofed customer requests (e.g., voice clones) may trigger unauthorised account changes or payments.

Regulatory officials

Fake directives from government agencies (e.g., urgent "audits") can pressure institutions to share sensitive data.

Recipients of deepfake attacks

Deepfake content is most often directed at:

Employees in finance or customer-facing roles

Staff with transaction authority (e.g., treasury teams) or access to sensitive systems.

External partners

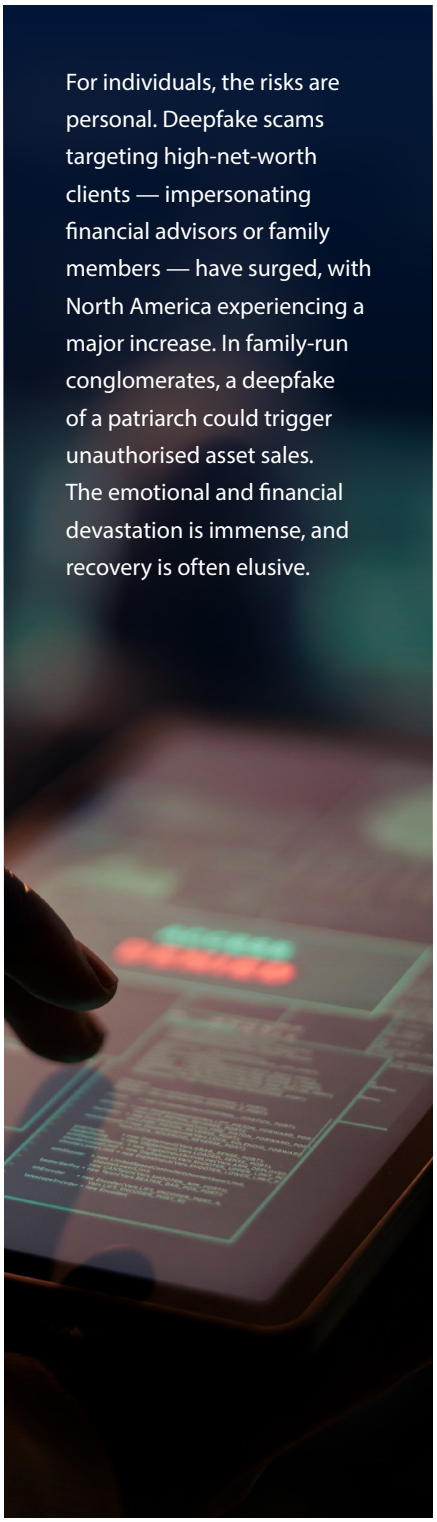
Vendors, auditors, or law firms tricked into sharing confidential data or approving fraudulent invoices.

Automated systems

Voice-based authentication tools or chatbots manipulated by AI-generated audio or text.

Customers

Retail clients targeted via fake investment ads, phishing calls, or fraudulent account recovery requests.



Financial fraud prevention: Technological and strategic measures to combat deepfake threats

As deepfake technology becomes more sophisticated, organisations must adopt a multi-layered approach to mitigate risks and safeguard sensitive information. While no single solution can fully eliminate the threat, a combination of cybersecurity measures, policy interventions, and advanced verification techniques can significantly reduce the impact of deepfake manipulation.

Strengthening cybersecurity culture and practices

One of the first lines of defence against deepfake-related threats is fostering a robust cybersecurity culture across all corporate activities. Employees are often the weakest link in security, making awareness and training essential components of any counter-deepfake strategy.

Education and awareness

Employees should be trained to recognise deepfake-generated content, especially in emails, video calls, and social media interactions.

Human firewalls

Encouraging proactive behaviour by employees, such as reporting suspicious content and verifying media authenticity before acting, is crucial in preventing financial and reputational damage.

Enterprise-wide cybersecurity policies

Organisations must implement strict guidelines regarding the verification of digital content before it is shared or used in decision-making. This includes checking the authenticity of video calls, email communications, and multimedia files.

Policy efforts: Embedding traceability and watermarks

Policy interventions play a key role in tackling deepfake threats by ensuring that AI-generated content remains traceable and identifiable. Regulatory bodies and corporate policies should enforce accountability in the creation and distribution of deepfake media.

Mandating watermarking and metadata tracking

Generative AI and Large Language Model (LLM) providers must embed digital watermarks into AI-generated content, allowing it to be easily detected and traced back to its source. By embedding cryptographic signatures, organisations can identify altered content and verify its authenticity.

AI-governed content authentication

Companies should integrate AI-driven authentication systems that scan videos and images in real time to detect manipulated elements. These systems can be used across social media platforms, financial institutions, and internal communications to prevent deepfake fraud.

Legislative action and industry collaboration

Governments and private organisations need to work together to create standards that define ethical AI use and deepfake detection mechanisms. Implementing compliance frameworks can push AI providers to take greater responsibility in preventing misuse.

Enhancing identity verification to combat advanced forgeries

As deepfakes evolve, identity verification processes must become more advanced to prevent fraudulent activities such as impersonation scams and financial fraud. Traditional security measures like passwords are no longer sufficient, making biometric and behavioural verification essential.

Multi-Factor Authentication (MFA)

Organisations should implement layered authentication protocols, requiring users to verify their identity through multiple means, such as biometrics, SMS codes, and security questions.

Behavioural biometrics

AI-driven behavioural analysis can detect anomalies in voice, typing patterns, and facial movements to determine whether an individual is authentic or digitally manipulated.

Single Sign-On (SSO) with AI-based anomaly detection

By integrating AI with SSO solutions, companies can track unusual login attempts and flag suspicious activities that may indicate identity fraud.

Privileged Access Management (PAM)

Restricting access to sensitive systems can prevent deepfake-driven social engineering attacks, where fraudsters attempt to manipulate employees into granting unauthorised access.

Implementing security measures from the outset

Instead of reacting to deepfake-related threats, organisations should adopt proactive security measures by embedding AI-driven fraud detection systems and continuously assessing potential vulnerabilities.

Identifying critical risk indicators

Companies must continuously monitor their digital ecosystems for signs of deepfake infiltration, including sudden changes in user behaviour, unusual transaction patterns, and suspicious communications.

Assessing financial and security risks

Deepfake-related fraud can lead to severe financial losses, legal liabilities, and reputational damage. Conducting regular risk assessments can help businesses prepare for potential threats and minimise exposure to AI-generated deception.

AI-powered fraud detection

Machine learning models trained to detect deepfake content can scan media files in real time, flagging manipulated videos, audio, and images before they cause harm.

The future of deepfake detection

As deepfake detection tools improve, so will deepfake technology. Quantum computing could soon enable real-time rendering of hyper-realistic forgeries, while AI-generated text (e.g., ChatGPT) may automate social engineering at scale. Conversely, breakthroughs in explainable AI and digital watermarking offer hope for pre-emptive defence.

A combination of strong cybersecurity policies, AI-powered detection mechanisms, and regulatory oversight is crucial in mitigating the risks associated with deepfakes. By adopting proactive measures and fostering collaboration between businesses, governments, and technology providers, we can ensure a safer digital ecosystem where authenticity and trust remain intact.

For more information, contact infosysbpm@infosys.com

Infosys®
Navigate your next