AI Deepfake: *Anatomy of An Attack*

AI-generated deepfakes are on the rise globally, affecting businesses and individuals alike.

An **AI deepfake** is synthetic media — video, audio, or imagery — generated by artificial intelligence to convincingly replicate a real person's likeness, voice, or behaviour. In a financial context, deepfakes are used to impersonate senior executives in real-time calls, making them extremely difficult to distinguish from genuine communication. **Recognising each stage of the attack and knowing the right response is among the most important steps you can take to help reduce exposure and protect yourself.**

What You Need To Know

Deepfakes are not easy to detect visually

Modern AI deepfakes are extremely difficult to distinguish from genuine video and voice calls. Attackers use real-time tools that replicate executives convincingly. Visual cues alone cannot usually be relied upon for verification.

Approval processes alone are usually not sufficient

Attackers specifically engineer scenarios to bypass standard controls by creating urgency, secrecy, and pressure to make exceptions. Strong processes must be supported by a culture of verification.

The 5-Stage Attack Lifecycle and How to Respond

1. Stage 1: Reconnaissance and Targeting

How it works
The attacker uses publicly available tools and social media to identify targets, collect voice and video data, and map organisational structures.

What to do
Limit your organisation's online exposure. Increase staff awareness and prepare teams for potential impersonation attempts.

2. Stage 2: Impersonation and Social Engineering

How it works
The attacker uses highly convincing impersonation to exploit trust in authority figures, creating urgency to bypass scepticism.

What to do
Enforce policy so that no business is conducted outside formal corporate channels. Define clear escalation protocols for unusual requests.

3. Stage 3: Deepfake Manipulation

How it works
AI tools are used to replicate authority figures through live video, voice cloning, and fabricated messages.

What to do
Empower employees to **challenge** authority and **verify** identity through independent channels. Establish clear senior executive communication protocols.

4. Stage 4: Payment Request and Control Bypass CRITICAL STAGE

How it works
Pressure and secrecy are used to bypass approval chains and create exceptions to standard controls.

What to do
Establish a policy that no single individual has sole payment authority. Require independent out-of-band verification and maintain standard controls for all urgent payments including new beneficiary approvals.

5. Stage 5: Funds Dispersed AFTERMATH

How it works
Late detection allows funds to be moved rapidly. Recovery becomes extremely difficult once a transfer has been completed internationally.

What to do
Run regular same-day reconciliation. Report suspected fraud to your bank as soon as practicable and seek to ensure staff are trained on incident response procedures.

Reinforce Awareness and A Culture Of Vigilance Through Regular Employee Training

Additional Resources

Citi recommends clients review the following materials to further strengthen their fraud prevention posture.

Beneficiary Change Request
Controls and verification steps for managing beneficiary change requests securely.

What to Do in the Event of Fraud
Immediate steps and contacts if a fraudulent payment has occurred.

AI Deepfakes: When Seeing and Hearing Cannot Be Trusted
Expert analysis of the deepfake threat landscape and actions organisations should take.

26-1D144E5D 07/26