

Deepfake and Voice Cloning Verification Protocols

Type: Process Guide

Target Audience: Security Teams, Finance, Executive Assistants, IT

This guide establishes organizational defenses against AI-powered fraud and synthetic media attacks. Deepfakes and voice cloning can now convincingly impersonate executives, employees, and trusted contacts.

■ **CRITICAL:** Never authorize high-value transactions, sensitive data transfers, or credential changes based solely on voice/video communication—even if it appears to be from a known executive.

1. Multi-Factor Authentication for High-Value Actions

Require multiple verification factors for transactions or requests that could cause significant harm if fraudulent.

High-Value Transaction Thresholds

Action Type	Threshold	Required Verification
Wire transfers	> \$10,000	2-person approval + callback
Vendor payment changes	Any amount	Written request + callback to known number
Credential/access changes	Any privileged account	In-person or video + challenge code
Data export requests	Sensitive/bulk data	Manager approval + IT verification
Emergency requests	Any "urgent" executive request	Mandatory callback + challenge code

2. Challenge-Response Systems

Establish pre-arranged secret words or codes that cannot be known by an attacker, even one with access to public information or hacked communications.

Implementation Checklist

- Assign unique challenge codes to executives and key personnel
- Store codes securely (not in email, chat, or shared drives)
- Rotate codes quarterly or after any suspected compromise
- Train staff to request code verification for sensitive requests
- Document code verification in transaction records

Example Protocol: "Before processing this wire transfer, I need to verify with our security code. What is the response to 'Alpha'?" — Correct response confirms identity.

3. Second-Channel Callback Verification

Always verify sensitive requests through a separate communication channel. If contacted by phone, hang up and call back using a known, pre-verified number.

Callback Protocol

1. **Receive request** via phone/video/email claiming to be from executive or vendor
2. **Do NOT act** on the request immediately, regardless of stated urgency
3. **Hang up** the current call (do not use "hold" or transfer)
4. **Look up** the verified contact number from internal directory
5. **Call back** using the verified number to confirm the request
6. **Use challenge code** to verify identity
7. **Document** the verification in the transaction record

4. Content Provenance Standards

Adopt technical standards for verifying the authenticity and origin of media content.

C2PA (Coalition for Content Provenance and Authenticity)

- Implement C2PA-compliant tools for creating organizational media
- Train communications team on content signing procedures
- Verify C2PA credentials on received media when available
- Publish organizational signing certificates for external verification

Detection Tools

- Deploy deepfake detection software for high-risk communications
- Integrate detection into video conferencing platforms
- Establish baseline voice/video samples of executives for comparison
- Subscribe to threat intelligence feeds for emerging deepfake attacks

5. Rapid Response Crisis Team

Establish protocols for when a deepfake attack is detected or suspected.

Crisis Response Checklist

- IMMEDIATE: Halt all pending transactions related to the suspected deepfake
- IMMEDIATE: Alert Security, Legal, and Communications teams
- Preserve all evidence (recordings, emails, logs)
- Verify status of any completed transactions
- Contact financial institutions if funds were transferred
- Brief executives on the attack and implement additional verification
- Report to law enforcement if financial loss occurred
- Conduct post-incident review and update protocols

6. Employee Training Program

Train all employees to recognize signs of voice cloning and video manipulation.

Recognition Signs - Voice Cloning

- Unusual urgency or pressure to act immediately
- Requests to bypass normal approval processes
- Slight audio artifacts, robotic quality, or unnatural pauses
- Caller unable to answer personal/contextual questions
- Request for secrecy ("don't tell anyone about this")

Recognition Signs - Video Deepfakes

- Unnatural blinking patterns or eye movement
- Lip sync slightly off from audio
- Unusual lighting or skin texture
- Blurring around face edges or hair
- Person unwilling to turn head or move naturally

- Conduct annual deepfake awareness training for all staff
- Run simulated deepfake exercises for finance/executive teams
- Include deepfake scenarios in security awareness program