

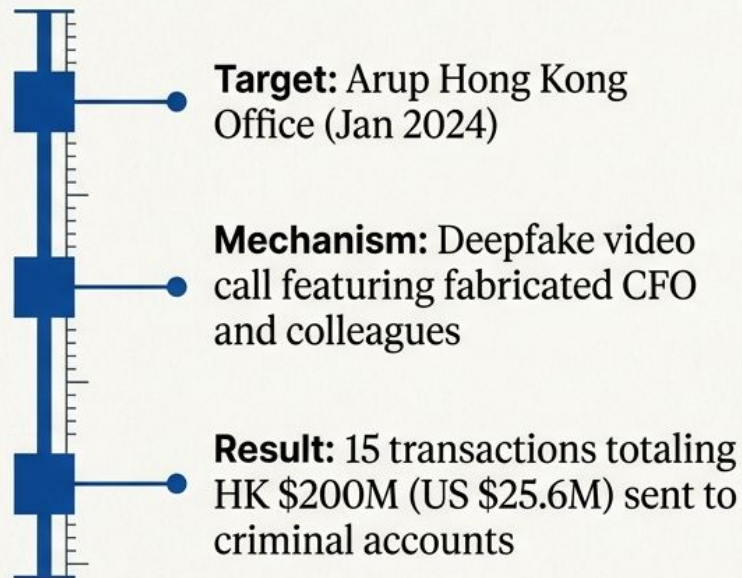
The Verification Blueprint

Defeating Executive Impersonation in the AI Era

Why deepfakes are an authority problem—and the 5 strategic pillars to secure your organization.

Based on research by
Jonathan H. Westover, PhD.

The Threat Reality: Anatomy of a US \$25.6M Loss



The psychological failure point:

The employee was skeptical.

But the apparent presence of senior leaders, the confidentiality of the request, and the implicit cost of questioning superiors overrode his doubt.

“Persuaded to do something they believed was legitimate.”

— Arup CIO (2025)

Escalating Threat Dynamics



The Raw Material

Attackers clone voices from keynote speeches, earnings calls, and podcasts.

Example: Wiz CEO's voice was successfully cloned from a single public conference talk.



The Ease of Deployment

45 Minutes

The time it took Arup's CIO to build a real-time deepfake of himself using off-the-shelf, open-source software.



The Financial Scale

\$40 Billion by 2027

Projected US fraud losses driven by Generative AI (Deloitte, 2024). AI-enabled Business Email Compromise already accounts for tens of millions in FBI complaints.

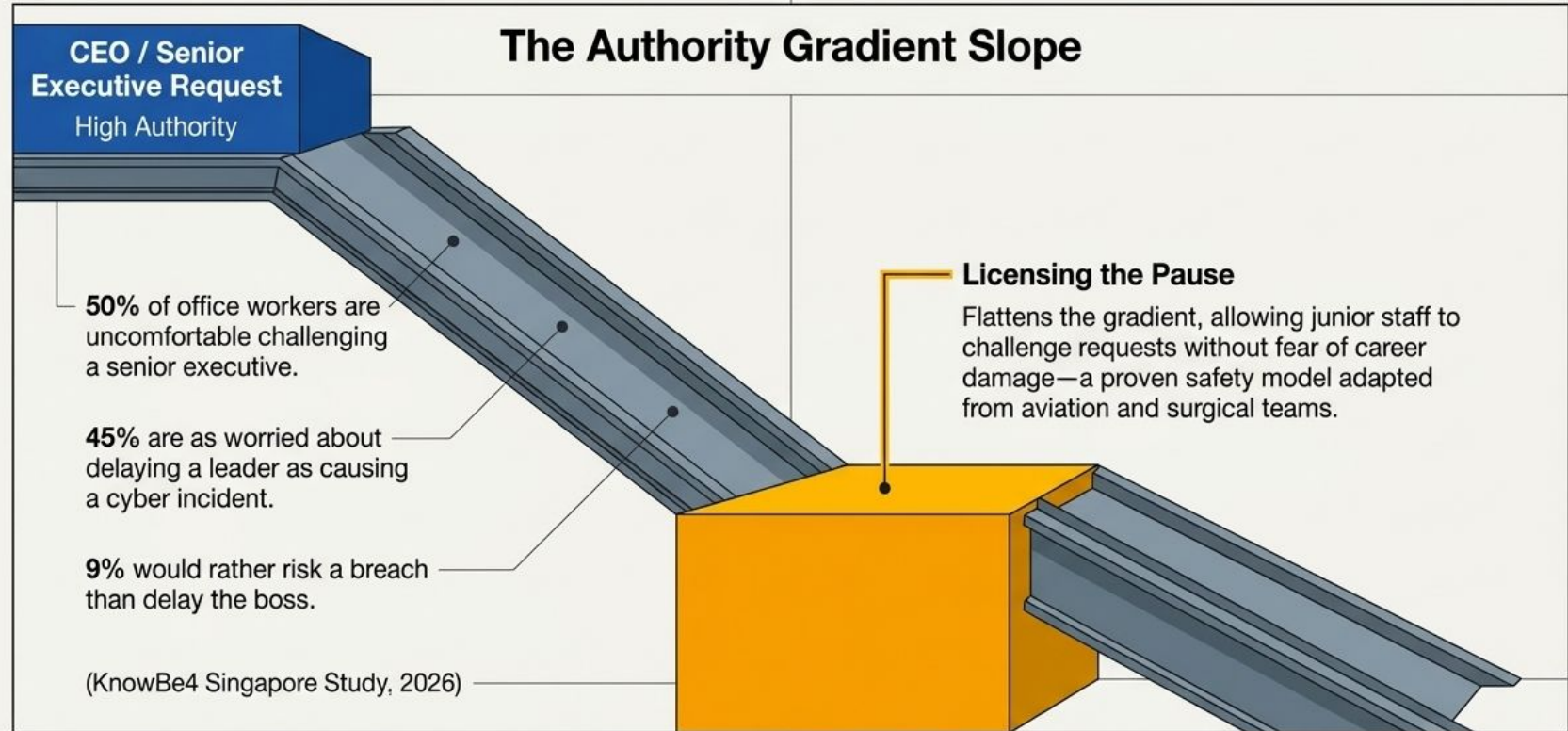
The Strategic Pivot: Changing the Paradigm

Diagnostic Comparison Matrix

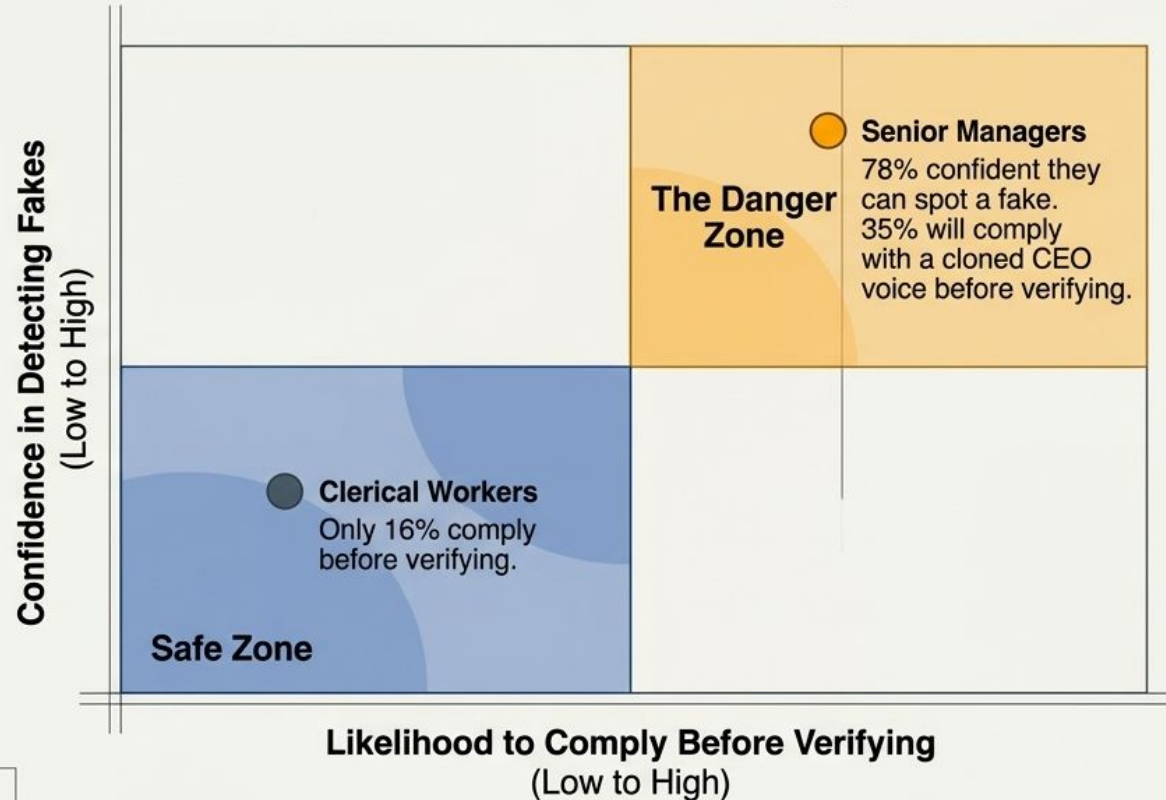
	The Old Paradigm	The Verification Paradigm
The Core Vulnerability	Technical failure (Our detection filters failed).	Authority failure (Our hierarchy was weaponized).
The Primary Defense Goal	Train employees to spot the fake (Look for visual glitches).	License the pause (Make it safe to delay and verify).
The Employee Action	Trust the video/voice if it seems authentic.	Treat voice/video as unverified by default; execute out-of-band call-backs.

Takeaway: We cannot rely on employees to spot the fake. We must build an architecture that expects deception.

The Psychological Vulnerability: The Authority Gradient



The Bias Blind Spot: Who is at Risk?



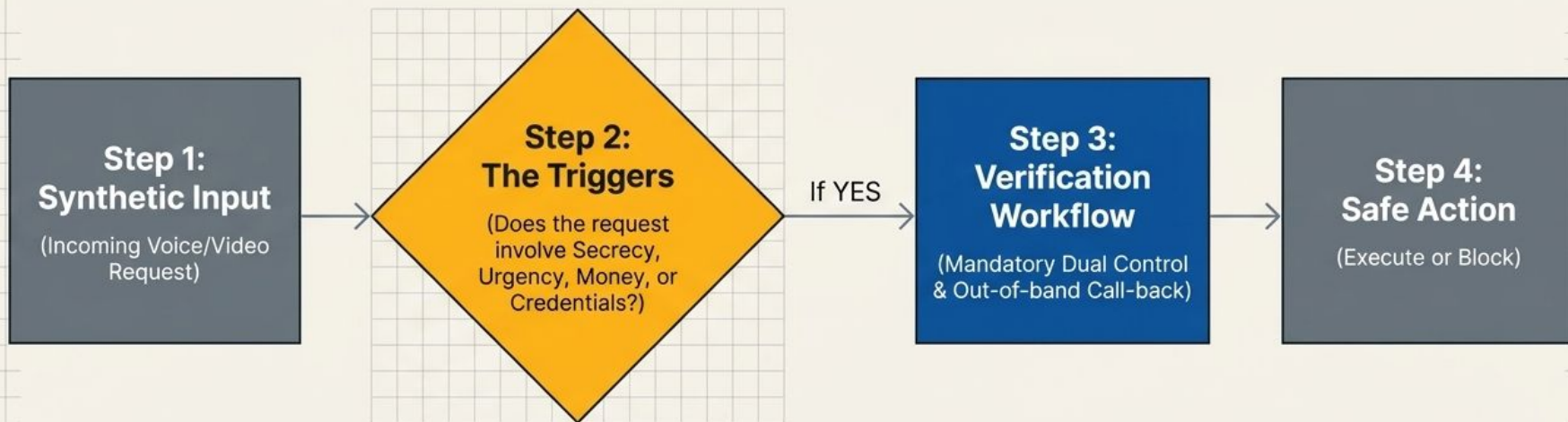
Key Insight

Overconfidence + proximity to power makes Senior Managers the ideal target.

Training that raises confidence without changing behavior worsens this specific risk.

The Anatomy of Interception

Moving from reliance on human detection to a structured verification architecture.



Playbook Step 1:

Leaders Who License the Pause



Explicit Permission: State publicly that verification is expected, never insubordinate.



Leaders Verify Too: Executives must visibly use the same call-back rules for their peers to normalize the behavior.



Praise the Catch: Publicly reward staff who delay a request to verify it, even if the request was legitimate.

Case Study: WPP & Wiz

After deepfake attacks, CEOs Mark Read (WPP) and Assaf Rappaport (Wiz) publicly shared the attempts, reminding staff that trusting an account photo or voice is a vulnerability—officially endorsing the pause.

Playbook Step 2: Out-of-Band Challenge Protocols

Core Principle: The question is no longer “Does this sound like the CEO?” but “Have I confirmed this on a channel the attacker does not control?”



Call Back on Known Numbers: Never use the number or link supplied in the urgent request.

Shared-Knowledge Challenges: Establish code phrases or personal questions a public AI model wouldn't know.

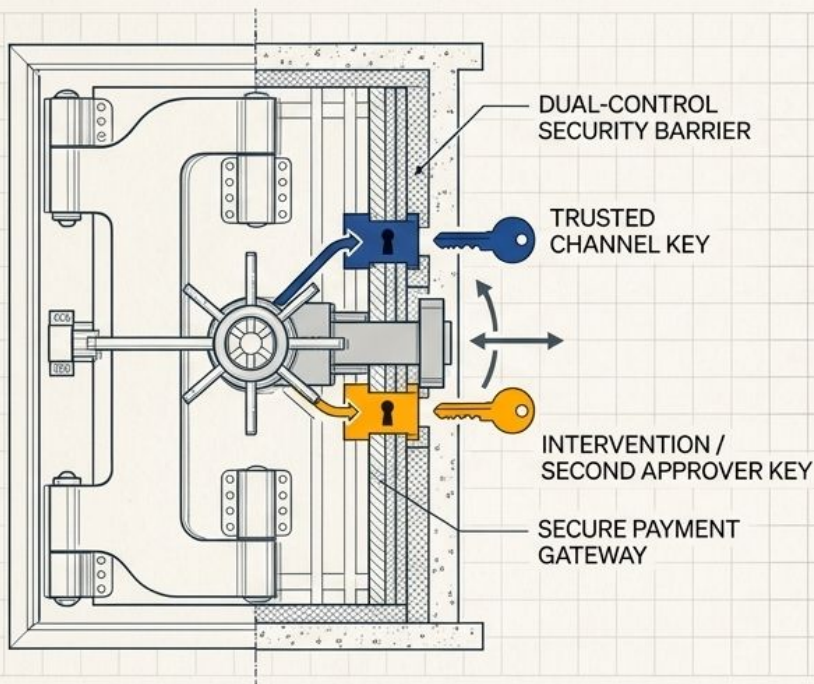
Channel Discipline: Treat personal messaging apps as unverified by default.

Case Study: Ferrari (July 2024)

An executive received a highly convincing deepfake call from “CEO Benedetto Vigna”. The executive asked for the title of a book Vigna had recently recommended. The attacker hung up immediately.

Playbook Step 3: Payment Architecture Independent of Detection

Core Principle: Design controls assuming someone WILL be fooled. No single person's judgment should be the last line of defense.



- ✓ **Mandatory Dual Control:** Independent approval required for high-value transfers, tied to out-of-band verification.
- ✓ **No Voice/Video Overrides:** Explicit policies stating that no executive (including the CEO) can bypass payment controls verbally.
- ✓ **Rapid Bank Contact:** Pre-arranged procedures to freeze funds within hours.

Case Study: Singapore Multinational

A finance director transferred \$499k after a deepfake Zoom call. A second request triggered him to contact the bank, successfully freezing the funds via Singapore/Hong Kong police cooperation.

Playbook Step 4: Hardening Identity Workflows



Core Principle: Deepfakes frequently target IT Help Desks and account recovery processes where staff are trained to be helpful to “stressed” executives.

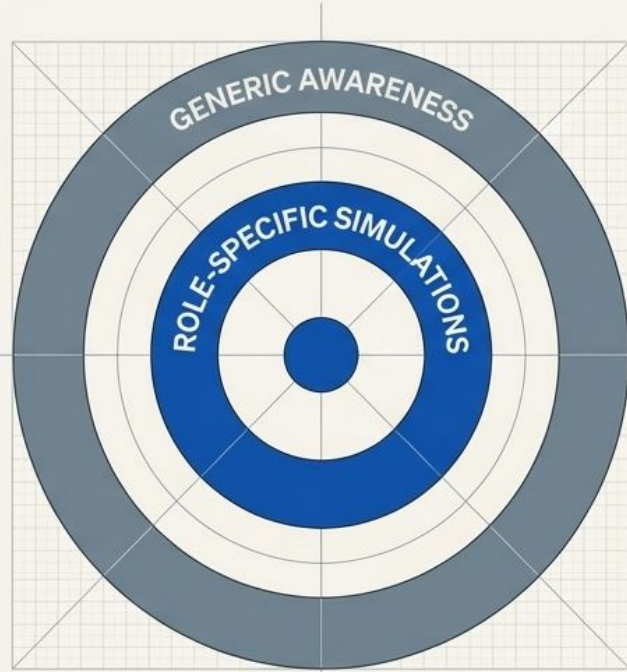
- ❖ **Phishing-Resistant Authentication:** Use hardware security keys or synced passkeys (NIST SP 800-63-4 standards) to eliminate shareable one-time codes.
- ✓ **Strict Help-Desk Checks:** Password resets require verifications that cannot be satisfied by voice alone.
- ✓ **Human Review for Sensitive Actions:** Privilege escalations demand independent secondary approval.

Case Study: Retool

An attacker used a deepfaked voice of a real IT staff member to **bypass MFA over a phone call**, resulting in the **complete takeover of 27 cloud customer accounts** in the cryptocurrency industry.

Playbook Step 5: Role-Specific Practice & Reporting

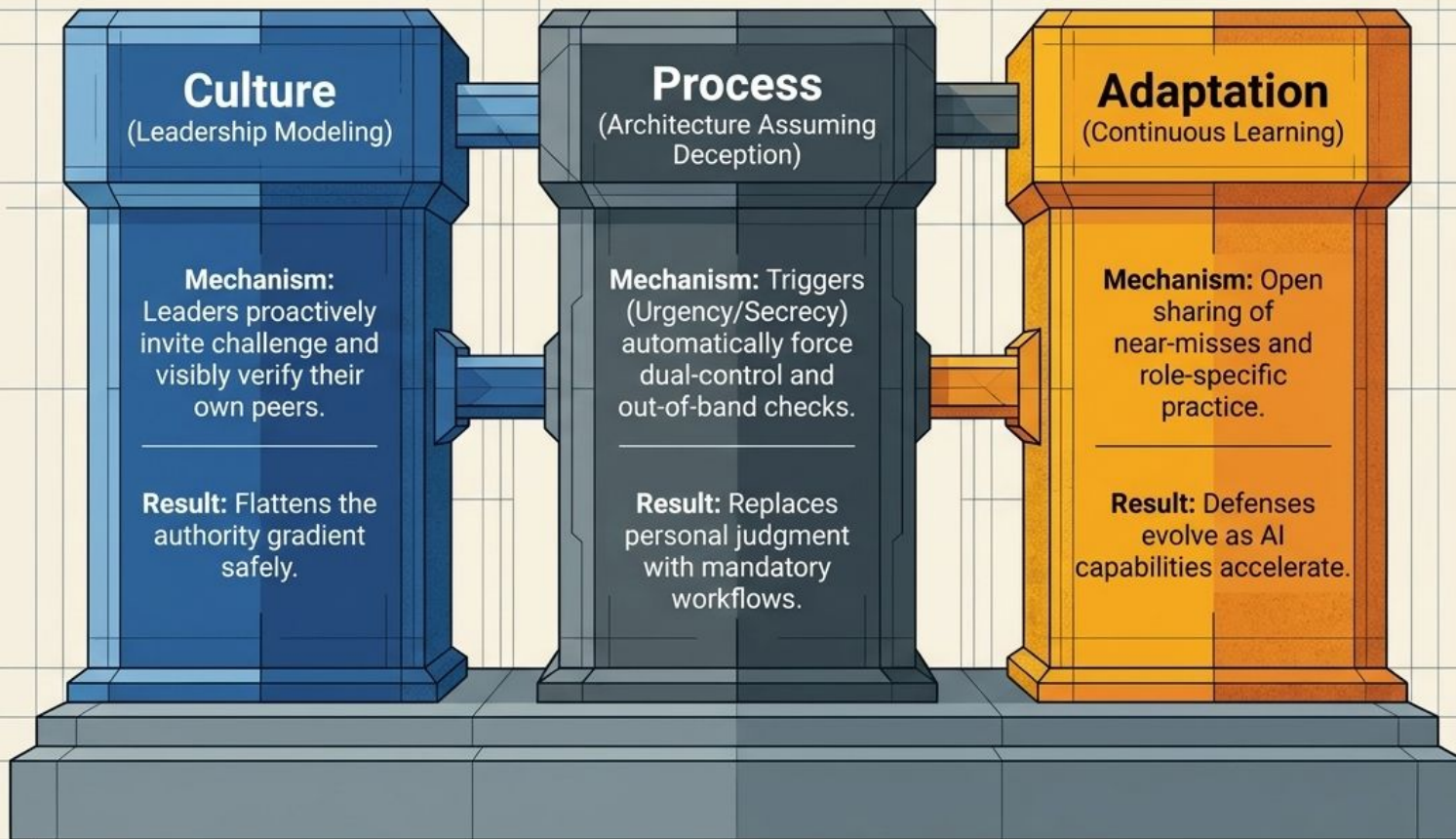
Core Principle: Generic “spot the typo” training fails. Training must address the authority gradient directly.



- ✓ **Target the Confident:** Focus live, interactive simulations specifically on Senior Managers, Finance Staff, and Executive Assistants.
- ✓ **Practice the Pushback:** Rehearse the exact scripts (“I’ll call you back on your usual line”) in real-time role-play.
- ✓ **Blame-Free Reporting:** Make reporting near-misses frictionless and publicly rewarded.

The Data: Research at UC San Diego Health (19,500 employees) showed standard annual awareness training had no meaningful relationship to stopping sophisticated attacks. Behavior, not awareness, is the metric.

Building Long-Term Verification Capability



“ No single convincing voice can move our money or open our systems.”

■ Verification is an expectation, not an accusation.

■ Channels dictate trust, not appearances.

■ Controls bind the C-Suite identically to the frontline.