



AI & Deepfakes: When Seeing Is No Longer Believing

WHAT ARE DEEPFAKES?



Deepfakes are AI-generated or manipulated content that can make it look or sound like someone said or did something they never did.

THEY CAN TAKE MANY FORMS



**FAKE
VIDEOS**



**SYNTHETIC
AUDIO**

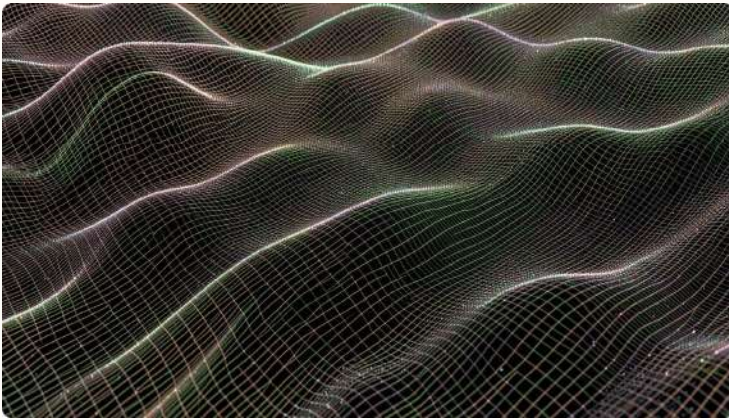


**ALTERED
IMAGES**



**FABRICATED
STORIES**

\$25.6 Million Vanished in a Single Video Call



REAL-WORLD CASE — FEBRUARY 2024

Arup, Hong Kong

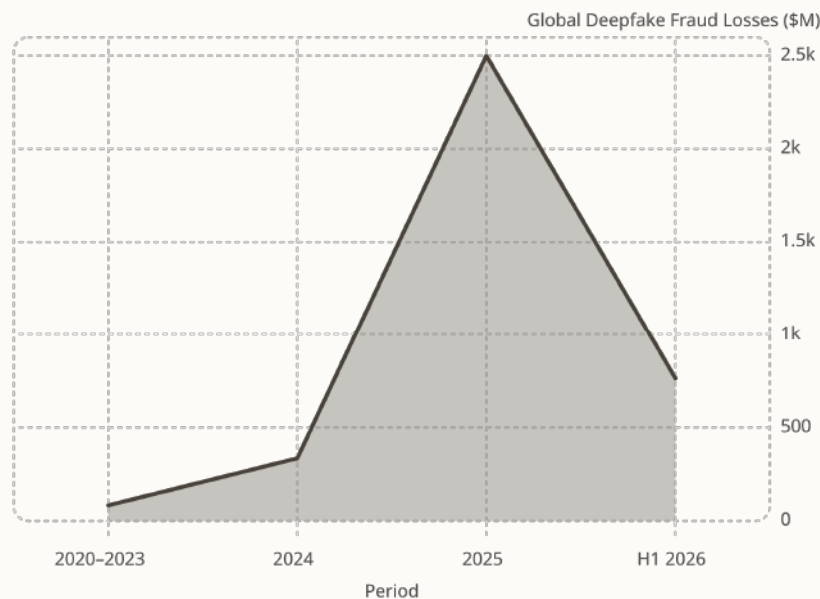
An employee joined what appeared to be a routine video call with the company's CFO and several senior executives. Every face. Every voice. Perfectly real.

All of them were AI-generated.

Convinced the meeting was legitimate, the employee transferred **\$25.6 million** across multiple transactions to fraudsters before anyone raised an alarm.

In a similar 2024 incident, **WPP** narrowly avoided an identical attack — a fake Microsoft Teams meeting using voice cloning — thanks only to employee suspicion.

\$3.7 Billion Lost — and the Crisis Is Just Beginning



The Numbers Are Staggering

Recorded global deepfake fraud losses have now reached **\$3.7 billion** — and are accelerating at a pace that outstrips nearly every other category of cybercrime.

Every 5 Minutes

A deepfake identity attack was recorded globally in 2024.

49% of Businesses

Encountered video deepfake fraud in 2024 — up from just 29% in 2022.

How Deepfakes Are Made

Generative Adversarial Networks (GANs)

Two AI models compete — one generates fake media, the other tries to detect it — cycling until the fake is **indistinguishable from real**.

Face Swapping & Voice Cloning

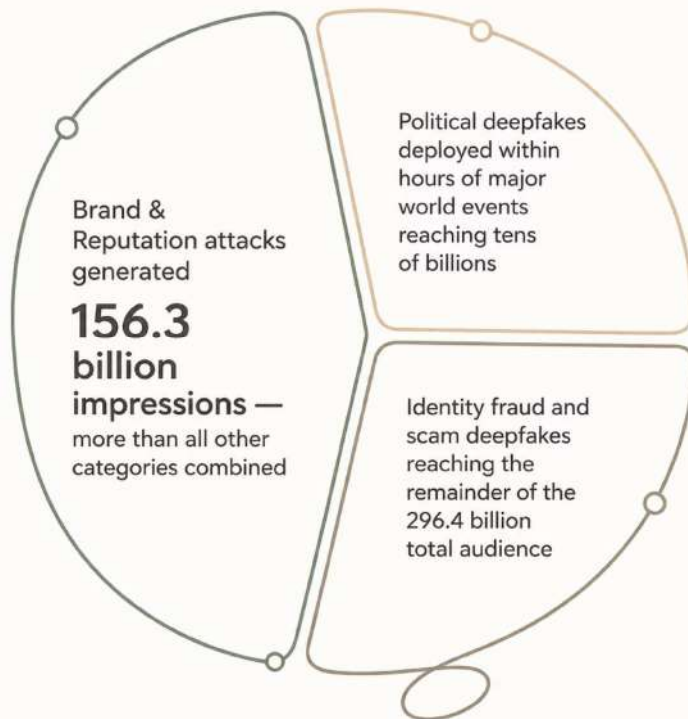
Trained on thousands of images and audio samples to perfectly replicate expressions, lip sync, and vocal tone.

Accessible to Anyone

Apps like FaceApp and open-source frameworks now let **non-experts** create convincing fakes in minutes — no coding required.



Deepfakes in Politics & Media



A Reach Beyond Imagination

In 2025, deepfake incidents collectively reached an audience of **296.4 billion** — roughly **37 times** the entire global internet population.

Brand Destruction at Scale

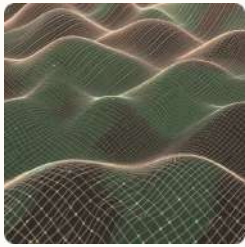
Brand & reputation attacks alone generated **156.3 billion impressions** — more than every other deepfake category combined.

Political Weaponization

Fabricated speeches and fake crises deployed within hours of major world events — before fact-checkers can respond.



Who Gets Targeted Most



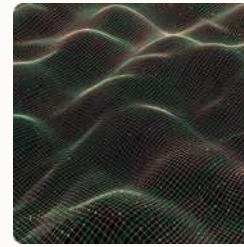
Women — Targeted 4.5× More

Female targets outnumber male targets by a 4.5:1 ratio — predominantly through non-consensual intimate imagery (NCII), a devastating and growing harm.



Celebrities & Executives

48% of Q3 2025 incidents targeted public figures — for impersonation scams, reputational attacks, and financial fraud.



Ordinary People

Identity fraud, romance scams, and financial manipulation are increasingly targeting everyday individuals — not just the famous or powerful.



How to Spot a Deepfake

Warning Signs to Look For

- 1

Visual Clues
Unnatural blinking, blurry edges around hair and ears, inconsistent lighting on the face — especially around the jawline.
- 2

Audio Clues
Robotic cadence, lip sync that doesn't quite match, strange background audio artifacts or sudden tonal shifts.
- 3

Context Clues
Urgent, unexpected requests — especially involving money transfers or credential sharing — are a red flag regardless of who appears to be asking.

AI Detection Tools

Purpose-built tools analyze metadata, pixel patterns, and audio inconsistencies in real time:

- ☐ **Best Practice:** Always verify through a second, independent channel before acting on any urgent request — no matter how convincing the video or voice.



Intel FakeCatcher
Real-time pixel-level analysis

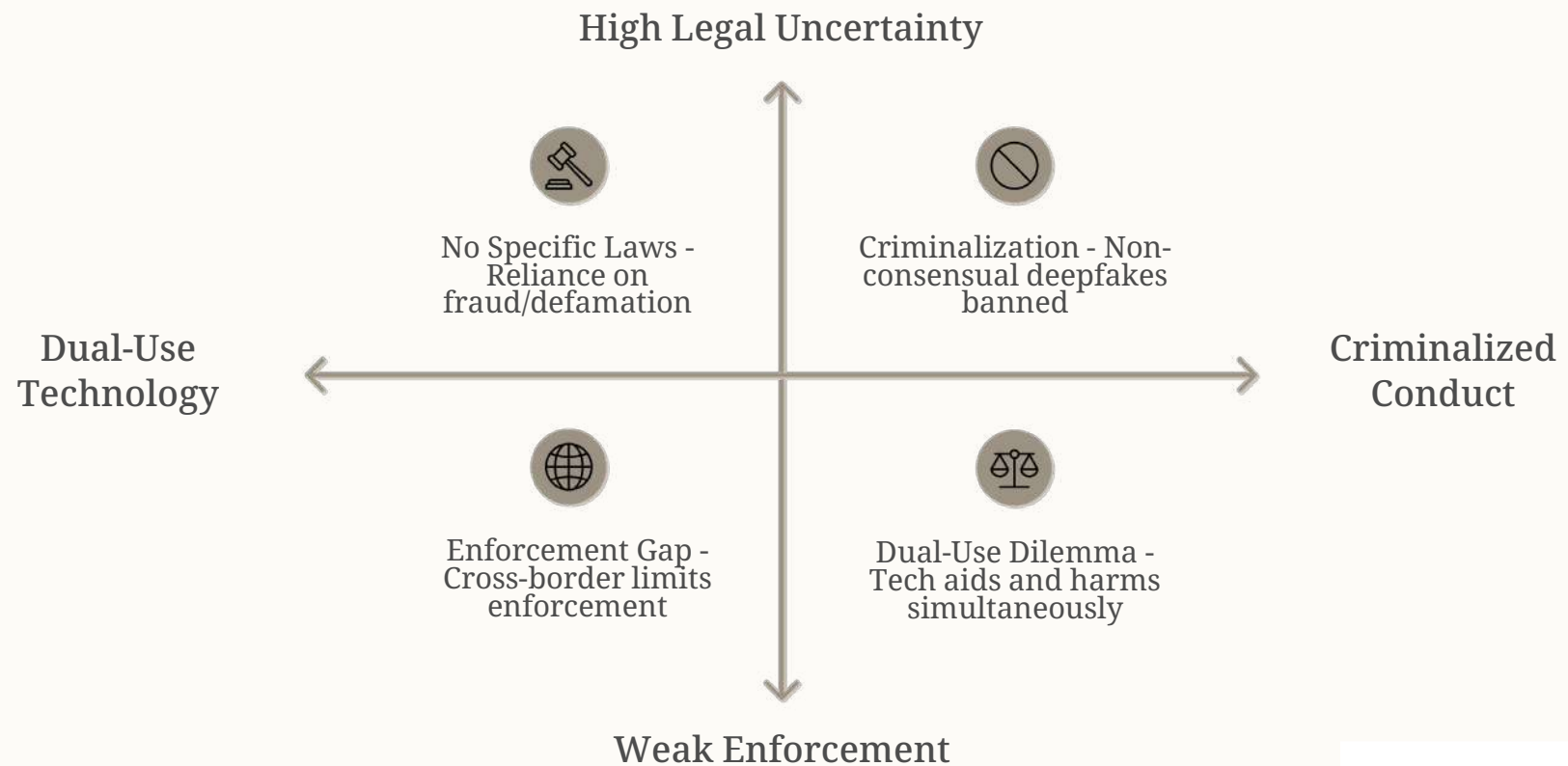


Resemble DETECT-2B
Advanced audio deepfake detection



Deepware Scanner
Free video analysis for everyone

The Legal & Ethical Landscape



The legal framework is catching up — but slowly. The central ethical question remains unanswered: **Who bears responsibility — the creator, the platform, or the AI tool provider?**



The Societal Stakes

The "Liar's Dividend"

Perhaps the most insidious effect: bad actors can now deny *real* evidence by simply claiming it's a deepfake — eroding trust in authentic footage and journalism.

Democracy at Risk

Fabricated political videos can swing elections and incite real-world conflict — often going viral in hours, long before fact-checkers can respond.

Financial System Vulnerability

85% of IT and cybersecurity professionals report their organization experienced a deepfake incident in the past 12 months. The average loss: **\$280,000 per attack.**

What You Can Do Right Now



Pause Before Sharing

Reverse-image search suspicious content, trace it to original sources, and look for corroborating reports from credible outlets before passing it on.



Demand Platform Accountability

Support and advocate for policies requiring **clear AI-generated content labels** on social media platforms — the public has a right to know what's real.



Use Detection Tools

Free tools like **Deepware Scanner** can flag suspicious videos in seconds. Make it a habit before trusting any unexpected video or audio message.



Stay Informed

Deepfake technology evolves weekly. **Digital literacy is your strongest defense.** The more you understand the threat, the harder you are to deceive.

WHAT ARE DEEPFAKES?



Deepfakes are AI-generated or manipulated content that can make it look or sound like someone said or did something they never did.

THEY CAN TAKE MANY FORMS



**FAKE
VIDEOS**



**SYNTHETIC
AUDIO**



**ALTERED
IMAGES**



**FABRICATED
STORIES**

WHY IT MATTERS



Erodes trust in media, institutions and people



Spreads false narratives and incites confusion



Harms reputations and privacy



Can influence markets and public opinion



Threatens security and social harmony

HOW TO STAY SAFE



Verify the source before you believe or share anything.



Look for inconsistencies in visuals, audio and context.



Be skeptical of sensational or emotional content.



Use fact-checking platforms and reverse image search.



Promote digital literacy and spread awareness.



Sources

- [\\$3.7B lost to deepfakes, social media is the primary origin — Surfshark Research](#)
- [Deepfake Fraud Incidents: 2024 Statistics & Trends — FluxForce](#)
- [How Does Deepfake Detection Work — Resemble AI](#)
- [Deconstructing Deepfakes—How do they work and what are the risks? — U.S. GAO](#)
- [Deepfakes and Fraud: Real-World Examples of AI Misuse — DigitalCommons@UNO / NCITE](#)
- [2025 Deepfake Threat Report — Resemble AI](#)
- [Q3 2025 Deepfake Incident Report](#)
- [Beyond Detection: The \\$280K Reality of Deepfake Attacks — IRONSCALES](#)