IDENTIFAI

Deepfake Report – Category: fraud

Fraud Vertical

Author	identifAI Intel Team
Period	2026
Date	August 31, 2026
Version	1.0

Legal Disclaimer & Terms of Use

1. **Nature of the Report and Institutional Intent**
   This report, titled Deepfake Report – Category: fraud, is published by identifAI–a specialized entity dedicated to the development of defensive technologies against synthetic media and deepfakes. This document is intended exclusively for informational, educational, and observational purposes. It serves as a technical "Observatory" on the current state of deepfake technology and its impact on the information ecosystem. Under no circumstances shall this report be construed as a legal accusation, a formal indictment, or an attempt to defame, offend, or damage the reputation of any individual, public figure, political entity, or nation mentioned herein. No statement contained in this report is intended to allege, nor should it be construed as alleging, that any named or referenced individual has engaged in unlawful, wrongful, or otherwise reprehensible conduct. All references to named individuals – whether public figures, political representatives, corporate executives, or private persons – describe exclusively the role of such individuals as subjects of observed synthetic media incidents, meaning as targets or subjects of third-party-generated deepfake content. Such references carry no implication of personal misconduct, criminal activity, or reputational fault on the part of the named individuals.

2. **Limitation of Liability; Accuracy and Methodology**
   While identifAI strives to ensure the highest degree of accuracy and technical rigor, the field of synthetic media is characterized by rapid evolution and inherent opacity. Consequently, the information, attributions, and analyses provided are offered on an "as-is" basis, without warranties of any kind, express or implied, as to the correctness, completeness, timeliness, or fitness for any particular purpose of the data presented. All quantitative data and categorical attributions contained herein are the output of a structured incident harvesting and normalization pipeline, incorporating LLM-assisted classification subsequently reviewed and validated by human subject-matter experts. Notwithstanding such review, AI-assisted classification processes may be subject to systematic or incidental error; accordingly, no representation is made that the data is free from inaccuracy or bias. Readers are directed to the Methodology section of this report of this report for a full description of the analytical framework, its assumptions, and its limitations. identifAI, its affiliates, officers, employees, and contributors shall not be held liable for any direct, indirect, incidental, special, or consequential damages arising from: (i) errors or omissions in the data; (ii) the interpretation or use of findings contained in this report; or (iii) any reliance placed on this report by any third party.

3. **Non-Adversarial Attribution; Public Interest Basis**
   Any mention of specific politicians, governmental bodies, corporate entities, media organizations or nations is based on technical forensic analysis, statistical aggregation and open-source intelligence (OSINT) gathered in accordance with the methodology described in the Methodology section of this report. Such mentions do not imply a judgment of intent, guilt, or culpability on the part of any mentioned party; do not constitute a claim that any named party created, disseminated, or endorsed any synthetic media content; are made solely to describe observed patterns of targeting, dissemination, and platform exploitation within the synthetic media threat landscape. The objective of this report is the analysis of technical trends and threat patterns for the purpose of advancing public awareness and organizational resilience. This report does not intervene in geopolitical discourse, partisan political activities, or any active legal proceeding. Publication on a matter of public interest: This report addresses matters of demonstrated public interest, including the integrity of democratic processes, financial security, and the protection of individuals from synthetic media abuse. The analysis and any related statements are made responsibly, on the basis of the evidence described herein, and following reasonable verification procedures consistent with responsible journalistic and research standards.

4. **AI Disclosure and Transparency**
   In accordance with identifAI's AI Disclosure Policy, and in compliance with applicable transparency obligations under Regulation (EU) 2024/1689 (EU Artificial Intelligence Act), the reader is hereby notified of the following:

   - Textual Content: The text within this report has been drafted with the assistance of Artificial Intelligence systems and subsequently reviewed and verified by human subject-matter experts to ensure technical consistency and factual accuracy. Human review constitutes the authoritative layer while AI-generated text

serves as a drafting aid only.

- Visual and Synthetic Media: All images, diagrams, or synthetic examples used within this report have been generated or modified via AI-based generative systems. These visual assets are employed exclusively for illustrative and didactic purposes to demonstrate the technologies analyzed by the Observatory. They do not constitute evidentiary material, do not purport to depict real events, and, unless expressly stated otherwise, do not represent or reproduce the likeness of any real person. They are fully compliant with identifAI's internal transparency and ethical standards.

5. **Privacy and Data Protection**

The processing of personal data of identifiable natural persons referenced in this report, including public figures, is carried out on the following legal bases:

- Legitimate interest and public interest (Art. 6(1)(e) and (f), Regulation (EU) 2016/679 "GDPR") given the manifest public interest in monitoring and countering synthetic media threats to democratic processes and financial integrity;
- Research, journalistic, and scientific expression purposes, as recognised under Art. 85 GDPR and applicable national implementing legislation.

identifAI processes personal data to the minimum extent necessary for the analytical purposes described herein. Named individuals may exercise their rights under applicable data protection law by contacting identifAI via its privacy policy.

For readers outside the European Union: identifAI observes data protection principles consistent with GDPR standards as a baseline across all jurisdictions of distribution.

6. **Intellectual Property and Permitted Use**

1 Executive Summary

This periodic report by identifAI outlines the strategic evolution of deepfake threats based on our global observations spanning from January 2020 to August 2026. For this analysis, an incident is defined as: an observed case of a deepfake being used in the real world, as reported by English-based media sources globally OR mention to deepfakes on social media posts. Over the reporting period, a vast volume of incidents was observed globally, reflecting the escalating frequency of synthetic media exploitation. Our targeted analysis focuses on the industrialisation of AI-enabled Business Email Compromise (BEC) and financial scams.

Key Finding

The shift from theoretical risk to multi -million-dollar damages—exemplified by complex impersonation attacks targeting C-level executives—demands immediate alignment with emerging frameworks like the 2024 FTC Impersonation Rule and Interpol directives.

The Industrialisation of Financial Fraud

The financial sector and corporate enterprises are currently navigating an unprecedented threat landscape. Criminal syndicates are leveraging "Fraud-as-a-Service" platforms to bypass Know Your Customer (KYC) systems and execute high-stakes social engineering at scale. Isolating the core media modalities used in these deepfake-related fraud incidents highlights the multimodal nature of modern attacks:

- *Video Synthesis*: 42.7%
- *Unspecified*: 24.0%
- *Audio-based Deception*: 21.9%
- *Image Synthesis*: 8.5%
- *Text Generation*: 2.9%

(Note: Identity fraud operations, such as automated KYC bypasses, act as the primary objective for many of these incidents and rely overwhelmingly on the visual synthesis modalities listed above.)

In January 2024, a finance employee at Arup, a multinational engineering firm, wired HK$200 million (c. US$25.6 million) to threat actors across fifteen transfers after interacting with an entire video conference of deepfaked colleagues, including a highly convincing clone of the firm's Chief Financial Officer. Furthermore, in early 2026, threat actors utilized document forgery and ID morphing to bypass video liveness verification at ABN AMRO Bank[1], successfully opening 46 fraudulent accounts before detection. As projected by major financial think tanks, generative AI could drive United States fraud losses to $40 billion by 2027, underscoring the severe economic consequences of synthetic impersonation.

Target Topologies and Geographic Distribution

Threat actors exhibit highly organized target selection. The target distribution for deepfake incidents classified as fraud reveals a strategic focus on Private Individuals (18.2%), Corporate Entities (15.8%), Public Figures (6.2%), and Business Executives (4.0%). A significant 55.8% of cases fall into a combined Other / Unspecified category, representing missing victim metadata or targets outside primary classifications.

Geographically, among the top targeted jurisdictions, the United States leads with 34.09% of incidents, followed by India (7.73%), the United Kingdom (6.59%), Hong Kong (2.71%), Australia (2.30%), and Brazil (1.78%). The remaining 44.80% of incidents are distributed across more than 90 other countries. These jurisdictions overlap heavily with major financial hubs, reflecting where attackers concentrate their efforts to maximize extortion and wire fraud payloads.

[1]Source: https://www.dutchnews.nl/2026/03/man-opened-46-bank-accounts-using-deepfakes-and-stolen-ids/

Ecosystem Exploitation and Collection Sources

Reconnaissance and deployment of these threats rely heavily on established digital ecosystems. A summary of fraud incidents discovered across major social networks in our observations demonstrates the following distribution:

1. **X**: 42.1%
2. **LinkedIn**: 14.5%
3. **Facebook**: 12.8%
4. **TikTok**: 12.5%
5. **YouTube**: 8.3%
6. **Other**: 9.8%

Looking at social media platforms for deepfake incidents, X leads at 42.1%, followed by LinkedIn (14.5%), Facebook (12.8%), TikTok (12.5%), and YouTube (8.3%). Attackers heavily leverage platforms like LinkedIn to scrape high-fidelity audio and video of senior leadership, which are subsequently fed into open-source generative models to craft highly convincing impersonation attacks.

Regulatory and Law Enforcement Response (2025–2026)

In response to these escalating, multi-million-dollar damages, global regulatory bodies and law enforcement agencies have rapidly mobilized. The FTC Impersonation Rule, finalized in 2024, empowers regulators to heavily penalize entities that utilize AI to falsely pose as businesses or government agencies. Similarly, the 2026 Global Fraud Summit in Vienna, coordinated by the United Nations Office on Drugs and Crime and Interpol, emphasized the critical need to combat industrialized, AI-driven scams originating from international fraud centers. Interpol's recent Cyber Threat Assessment for the Asia and South Pacific regions highlighted a 600% increase in deepfake-related discussions on forums and Telegram channels popular among Southeast Asian threat actors between February and June 2024[2]. Security leaders must urgently adapt to this reality, moving beyond legacy verification protocols and implementing multi-layered cryptographic media authentication alongside continuous behavioral biometric defenses. For a detailed breakdown of regional mandates, see Section 7

[2] Source: https://industrialcyber.co/reports/asia-pacific-cyber-threat-environment-intensifies-as-interpol-records-surge-in-ransomware-phishing-ddos-attacks/

Contents

2 Key Findings

13.27% Observed Events — classified as fraud on overall classified events	**17.6%** Identity Fraud — of detected fraud events	**21.9%** Voice Cloning — of detected fraud events	**42.7%** Video Manipulation — of detected fraud events
59747 Total Observed Events	**7926** Total Fraud Events	**3384** Video Manipulations	**1805** Unique targets

Incident Definition Reminder: An observed case of a deepfake being used in the real world, as reported by English-based media sources globally or mentions to deepfakes on social media posts. Out of a total dataset spanning hundreds of thousands of documented incidents during the reporting period, this analysis deliberately narrows its focus to the subset specifically classified as fraud. By isolating these targeted cases from the broader dataset–which represents 100.0% of the overall threat volume evaluated–we conduct a granular deep dive into the underlying mechanics of these operations. Our findings highlight a critical pivot toward high-stakes financial deception, synthetic identity fraud, and advanced evasion techniques that represent the highest-risk segment of the observed threat landscape.

Key Finding

The United States was identified as the source or target of 34.09% of all observed threats, with India and the United Kingdom representing the next most targeted nations. Unclassified data aggregated within the broader 'Global' category was omitted from discussions.

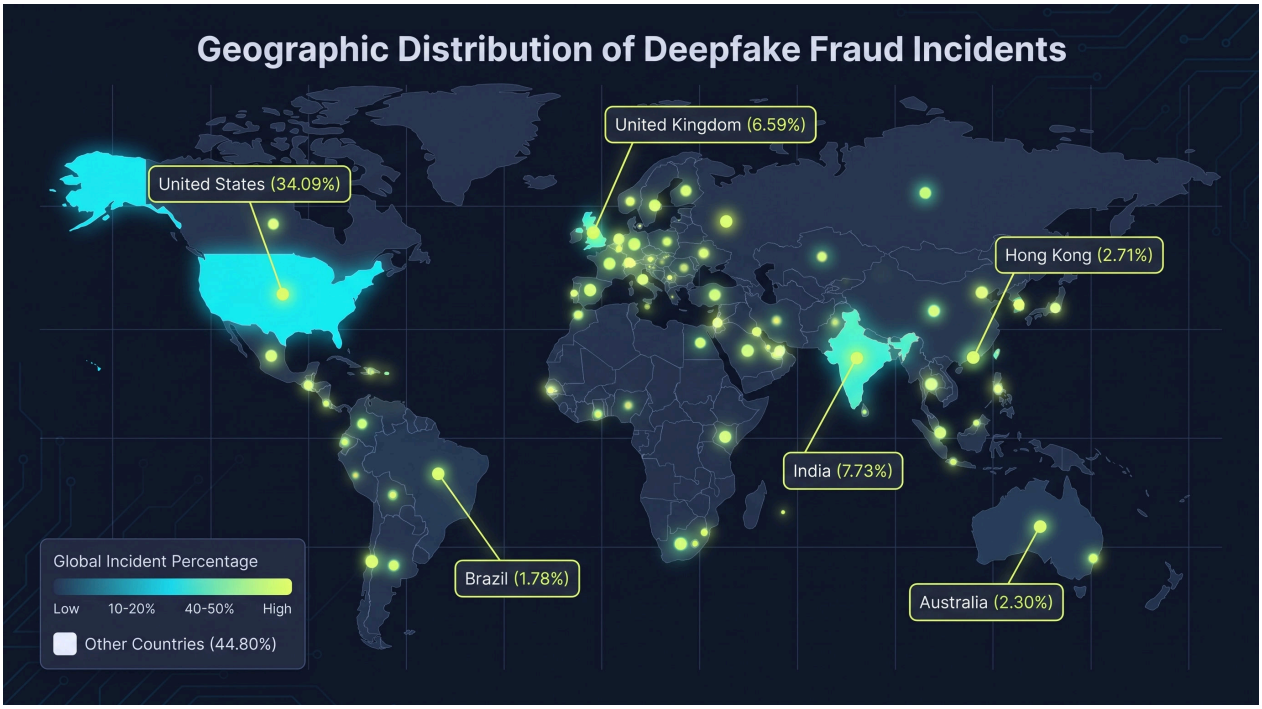


Figure 1: Global heatmap of deepfake fraud incidents by country, showing the concentration of threats in major financial and technological hubs. - AI generated image

Figure 1 illustrates the geographic distribution of these threats, underscoring a concentrated risk in major financial hubs. Under emerging frameworks such as European regulatory provisions and US FTC guidelines, multinational enterprises face intensifying compliance pressures. Regulatory bodies, including Europol and Interpol, emphasize that mitigating this globalized threat requires stringent liveness-detection and continuous biometric authentication in cross-border financial operations.

Key Finding

Video Synthesis dominates primary fraud modalities (42.7%, comprising Video Manipulation and GenAI Video), followed by Audio-based Deception at 21.9%.

The prominence of video and audio synthesis is directly linked to business email compromise and executive impersonation. During previous reporting cycles, a sophisticated attack on a multinational engineering firm successfully utilized deepfake video conferencing to impersonate corporate leadership, resulting in the fraudulent transfer of substantial corporate funds. By reconstructing executives' likenesses from publicly available media, threat actors execute full attack chains–from target selection to media synthesis and exploitation–with unprecedented precision.

Key Finding

X accounts for 42.1% of the platform exploitation footprint, with LinkedIn and Facebook serving as major amplification vectors at 14.5% and 12.8%, respectively.

The concentration of incidents on professional and public networks facilitates the social engineering necessary for corporate fraud. Threat actors systematically exploit these platforms by executing the following sequence:

1. Gathering open-source intelligence to identify finance workers and map enterprise hierarchies.
2. Training generative models on publicly available executive media to synthesize convincing audio and video.

Consequently, financial institutions are accelerating the deployment of session-level biometric countermeasures to defend against synthetic injections validated by manipulated social profiles.

Key Finding

Mainstream News Media drove 83.23% of global deepfake reporting, while Academic Journals accounted for 2.02%. These documented news reports represent approximately 8.7% of all classified fraud incident sources.

Public awareness of financial deepfakes remains heavily dependent on mainstream journalism, as seen in the extensive media visibility surrounding real-time face-swap attacks bypassing remote banking verification flows. A reported 77.0% of fraud professionals have noted noticeable increases in deepfake social engineering. Despite advanced AI capabilities, the detection of synthetic identities relies increasingly on passive signals–such as three-dimensional depth and blood flow–rather than traditional checks, compelling academic and corporate security researchers to continuously refine their defensive architectures.

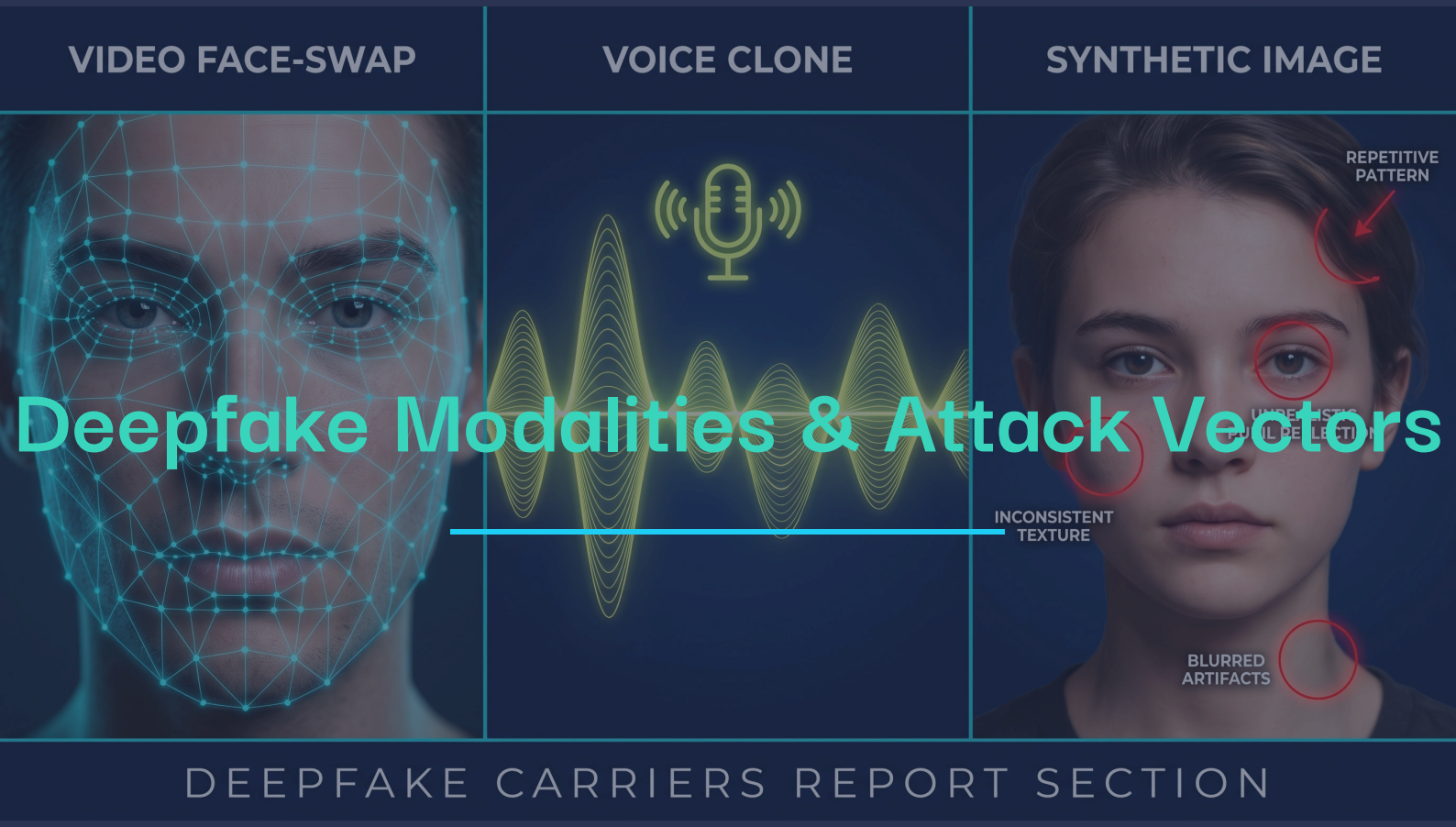
Key Finding

Among classified targets, Private Individuals account for 18.2% of fraud deepfakes, while Corporate Entities comprise 15.8%.

While broad public targeting persists, the synthesis of corporate assets presents the highest financial risk. Financial institutions are forced to pivot away from standalone identity verification models due to emerging attack vectors:

- Deployments of camera-injection tools bypassing traditional *Know Your Customer* verification infrastructure.
- Sophisticated cryptocurrency and investment fraud successfully exploiting corporate likenesses.

Defensive postures now demand multi-layered architectures capable of authenticating continuous biometric signals.


VIDEO FACE-SWAP
VOICE CLONE
SYNTHETIC IMAGE
REPETITIVE PATTERN
INCONSISTENT TEXTURE
BLURRED ARTIFACTS
Deepfake Modalities & Attack Vectors
DEEPFAKE CARRIERS REPORT SECTION

3 Attack TTPs & Modalities

Threat actors currently leverage a range of carrier modalities to execute sophisticated fraud campaigns, led by **Video Synthesis** (42.7%). This is followed by a significant **Unspecified** segment (24.0%)–which encompasses emerging modalities like generative text-to-video and cases where the carrier was not technically confirmed–**Audio-based Deception** (21.9%), **Image Synthesis** (8.5%), and **Text Generation** (2.9%). Meanwhile, 17.6% of the total fraud incidents were explicitly categorized by their primary objective: **Identity Fraud** (e.g., KYC bypasses).

In the observed fraud category incidents, attackers adhere to a rigorous fraud attack chain: *target selection → media synthesis → delivery channel → authorisation bypass → cash-out*. Reconnaissance for target selection relies overwhelmingly on open-source intelligence, with incident discovery mapped primarily to **X** (42.1%), **LinkedIn** (14.5%), and **Facebook** (12.8%). The targets of these synthetic media campaigns reflect a broad attack surface. While identified segments are led by **Private Individuals** (18.2%), **Corporate Entities** (15.8%), **Public Figures** (6.2%), and **Business Executives** (4.0%), a significant 55.8% of incidents fall into a combined **Other / Unspecified** category. This segment represents cases where reporting lacks specific victim metadata or the target falls outside the primary demographic classifications. Figure 3 depicts a conceptual mapping of the modular stages in a deepfake-enabled financial fraud operation.

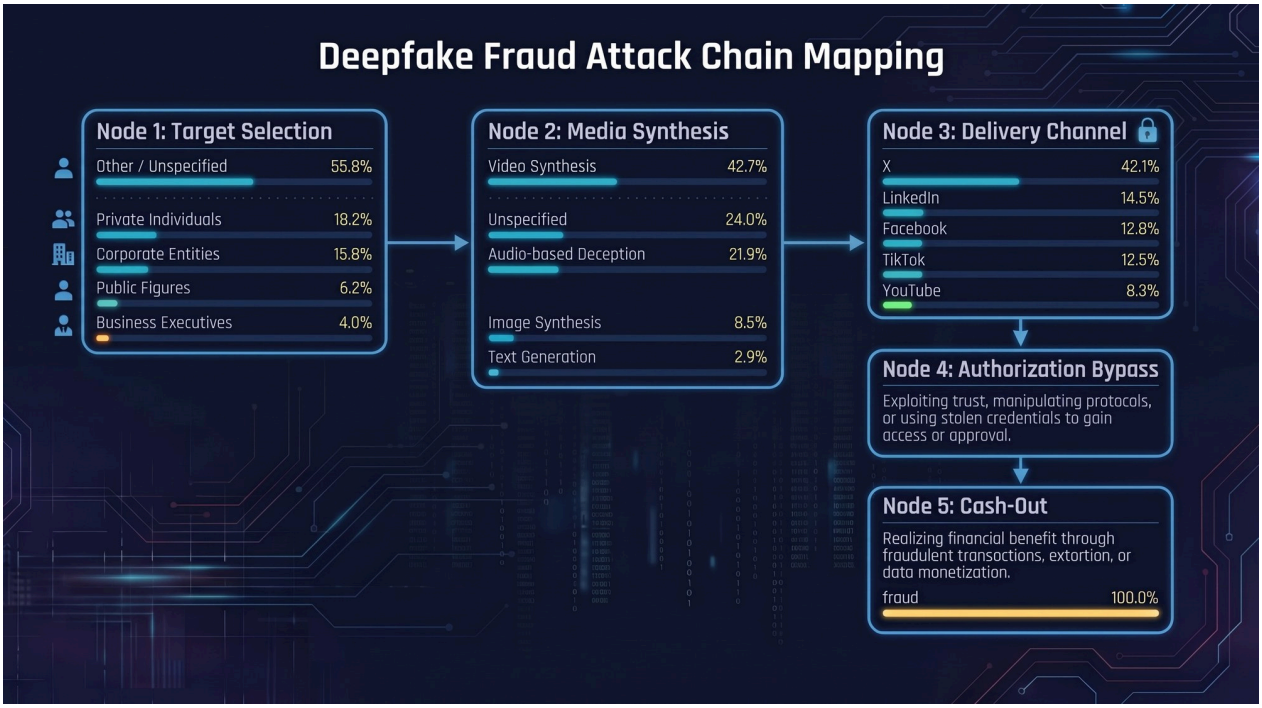


Figure 2: A conceptual mapping of the modular stages in a deepfake-enabled financial fraud operation. - AI generated image

Key Finding

Across our observations spanning January 2020 to August 2026, the percentage share of voice cloning surged month-by-month as open-source TTS barriers lowered. Threat actors are rapidly abandoning static phishing in favour of dynamic, multi-modal synthetic impersonation.

This temporal evolution has directly escalated the severity of Business Email Compromise (BEC) and executive impersonation. The efficacy of these carriers in bypassing payment-authorisation protocols was vividly demonstrated in January 2024

when a finance worker at Arup, a multinational engineering firm, transferred approximately $25.6 million after attending a live video conference. The employee was convinced by real-time video face-swaps and voice clones of a prominent CFO and several colleagues. The Europol Internet Organised Crime Threat Assessment (IOCTA) 2026 report explicitly highlights how this combination of real-time video and audio manipulation has become a staple for organised cybercrime syndicates targeting corporate wire-transfer mechanisms.

For banking and cryptocurrency platforms, the primary technical modality of concern is the deployment of deepfakes via camera injection techniques. Attackers exploit synthetic still imagery and generated video to bypass Know Your Customer (KYC) onboarding and biometric liveness detection. As noted in the World Economic Forum (WEF) Cybercrime Atlas 2026 report, commercial face-swapping tools and virtual-camera injection software are now commoditised in underground markets. These tools intercept and replace a device's physical webcam stream with a synthetic feed, allowing threat actors to project an AI-generated face over their own while presenting authentic or forged identity documents. Because the software perfectly replicates micro-movements, blink patterns, and lighting variations with sub-50-millisecond latency, it effortlessly defeats legacy passive biometric liveness checks.

Not long ago, manipulating digital video or audio required a Hollywood budget, a render farm, and a team of specialized VFX artists. Deception was expensive, slow, and computationally restrictive.

Today, that barrier to entry has completely evaporated. We are witnessing an unprecedented explosion in generative AI capabilities. According to recent cybersecurity benchmarks, the availability and creation of deepfake and AI-manipulated content have surged exponentially, growing by several hundred percent year-over-year.

The core issue isn't just the sheer volume; it is the simplicity. What once required deep technical expertise can now be executed by anyone with a smartphone and a basic web interface. The magnitude of this shift means that the threat vector is no longer elite nation-states or highly sophisticated cybercrime syndicates—it is virtually anyone.

Many risk compliance officers and product owners believe that because their KYC or insurance app uses advanced liveness detection or native camera APIs, it is immune to fraud. This is a profound misunderstanding of local hardware vulnerability.

Today, bad actors do not need to crack your app's source code or breach your cloud servers. Instead, they leverage low-cost, readily available tools—often costing just a few dozen dollars—designed to bypass local device environments entirely.

We are rapidly moving into a world dominated by Agentic AI—autonomous AI agents capable of navigating interfaces, mimicking human behaviors, and manipulating workflows at scale. In this new landscape, relying on local app-level isolation for security is an entirely obsolete strategy.

If your security model relies on the assumption that the user's device is pristine and uncompromised, your model is already broken.

The industry must adopt a fundamental paradigm shift: Assume the device is compromised. Always.

Just as the corporate network architecture transitioned from perimeter security to a "Zero Trust" model, identity verification and claims management must transition to a Zero-Trust Media environment. We can no longer trust the container (the app); we must ruthlessly validate the mathematical and structural integrity of the content itself.

Likely, regulatory bodies are aggressively adapting to this synthetic fraud wave. The US Federal Trade Commission (FTC) Trade Regulation Rule on Impersonation of Government and Businesses (16 CFR Part 461), finalized in 2024, specifically targets the commercial misuse of brand and government likenesses, allowing the FTC to seek direct civil penalties per violation. However, as the technical barrier to entry for media synthesis continues to collapse, financial institutions and enterprises must pivot from reactive detection to continuous, cryptographically verified authentication models to secure the delivery and authorisation phases of the modern fraud kill chain. As detailed in Section 7, regulators are increasingly classifying unmitigated camera-injection risks as operational compliance breaches under RBI, HKMA, and FFIEC guidelines.

4 Examples of Current Toolsets for Voice and Video Manipulation

The contemporary landscape of synthetic media generation is characterized by a dual-track proliferation of highly accessible open-source repositories and sophisticated commercial platforms, fundamentally altering the barrier to entry for generating high-fidelity deepfakes. On the open-source front, developer communities on GitHub continuously refine and democratize advanced models for both audio and visual manipulation.

For voice cloning, frameworks like the community-maintained *idiap/coqui-ai-TTS* (utilizing models like XTTS v2) offer fully localized, multi-lingual text-to-speech engines capable of zero-shot voice cloning with just a few seconds of reference audio. The barrier to creating a localized clone has been reduced to basic scripting. A standard implementation requires only a few lines of Python code to clone a target's voice entirely offline:

```python
from TTS.api import TTS

# Initialize the XTTS v2 model for zero-shot cloning
tts = TTS("tts_models/multilingual/multi-dataset/xtts_v2", gpu=True)

# Clone a voice from a short audio sample and generate new speech
tts.tts_to_file(
    text="This audio is entirely synthetic and cloned from a short sample.",
    speaker_wav="target_victim_voice.wav",
    language="en",
    file_path="cloned_output.wav"
)
```

Listing 1: Zero-shot voice cloning using Coqui TTS API

Concurrently, video manipulation and face-swapping algorithms have achieved unprecedented photorealism through active open-source projects. Repositories like *ai-forever/ghost*, alongside highly popular community-driven tools such as *FaceFusion*, *Roop*, and *Wav2Lip*, empower users to seamlessly blend source identities into target footage or synchronize lip movements to fabricated audio tracks. Often, these sophisticated video manipulations abstract away the underlying neural network complexities, allowing execution via streamlined command-line interfaces:

```
python run.py \
    --source attacker_selected_face.jpg \
    --target original_CEO_video.mp4 \
    --output deepfake_video.mp4 \
    --execution-providers cuda
```

Listing 2: One-shot face swap via FaceFusion CLI

Parallel to this decentralized ecosystem, professional commercial services have productized these generative models into frictionless, API-driven workflows. Platforms such as *ElevenLabs* currently dominate the synthetic audio domain. They offer proprietary voice changers, instant voice cloning, and ultra-low-latency voice agents that deliver studio-grade vocal mimicry via simple REST endpoints or official Python SDKs.

In the visual domain, commercial enterprise solutions like *HeyGen* and *Synthesia* provide comprehensive avatar animation suites, allowing users to transform a single high-resolution image into a lifelike, articulating character perfectly lip-synced to an AI-generated voiceover. This powerful convergence of free, barrierless GitHub repositories and subscription-based, high-fidelity commercial platforms has created a robust toolset that arms both legitimate digital creators and malicious actors with an unprecedented capacity to programmatically produce highly deceptive, multi-modal synthetic media.

Beyond these enterprise-grade services, the consumer market has been rapidly saturated with hyper-accessible, low-cost applications designed specifically for identity manipulation. Tools such as DeepFaceLive and the FaceSwap App have effectively democratized both real-time and post-processed face replacement. DeepFaceLive, for instance, enables users to swap faces on live video streams using standard consumer GPUs, making it incredibly easy to execute live social engineering or impersonation attacks during video calls.

Simultaneously, mobile-centric platforms like FaceSwap bypass the need for expensive hardware entirely by leveraging cloud processing. For just a few dollars a month in subscription fees, these apps offer intuitive, one-click interfaces that abstract away all the underlying machine learning complexity. This severe commoditization means that producing a

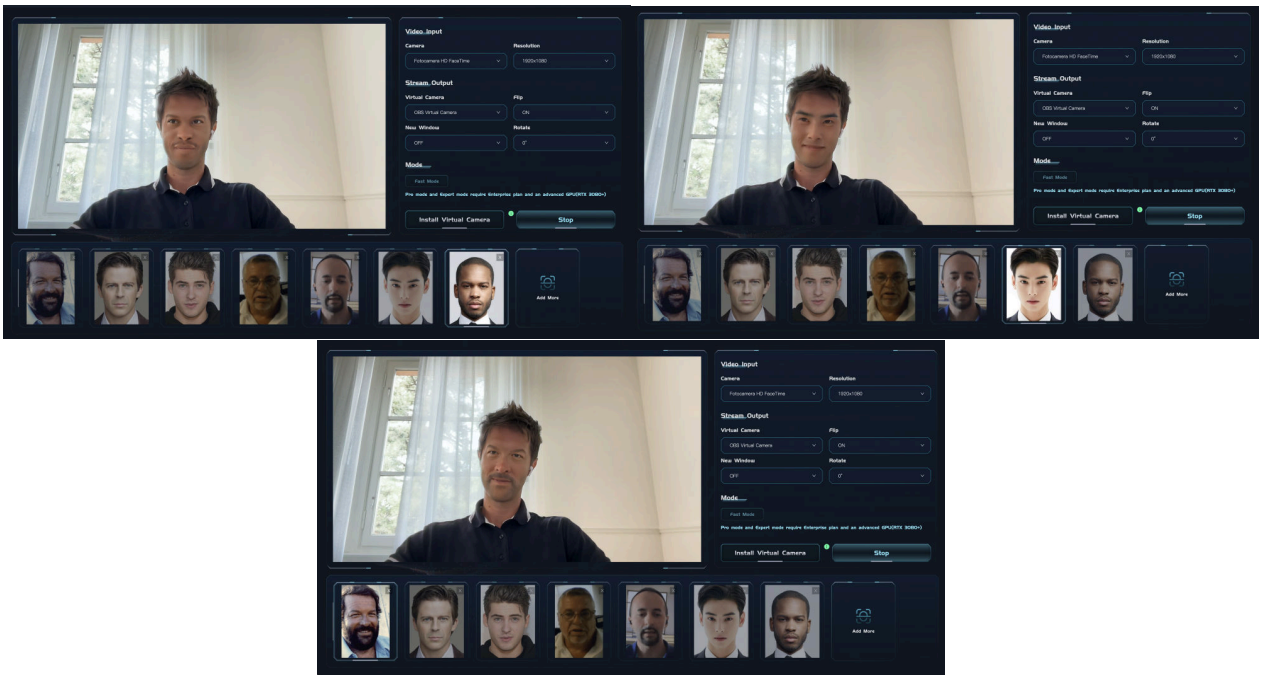


Figure 3: Examples of faceswaps using SwapFace App.

convincing deepfake no longer requires thousands of dollars in computing power or a background in computer science. Instead, high-quality visual spoofing can be achieved by virtually anyone for the cost of a cup of coffee, drastically lowering the barrier to entry and amplifying the scale at which synthetic deception can be deployed.

5 Social Platform Exploitation

Threat actors systematically exploit social media infrastructure, transforming communication networks into high-volume delivery mechanisms for synthetic fraud. Based on our observations of incidents classified entirely under fraud, attackers weaponize these architectures to execute sophisticated corporate impersonation and financial scams.

Key Finding

Social media platforms are the primary delivery vectors for deepfake fraud, with X accounting for 42.1% of observed distribution, effectively capturing victims through viral reach and algorithmic amplification.

Fraud Distribution Channels The distribution of deepfake-enabled fraud is increasingly dominated by platform-specific amplification. As shown in Figure 4, the primary conduits for these campaigns are:

- **X (42.1%):** The platform remains the primary vector for the viral spread of synthetic media. Attackers exploit the real-time nature of the service and the effectiveness of automated bot swarms to amplify deepfake videos of public figures, often bypassing community notes through rapid-fire account rotation.
- **LinkedIn (14.5%):** Exploitation here focuses on high-value corporate fraud. By distributing synthetic media via professional networking channels, attackers leverage the implicit trust of the platform to execute executive impersonation scams and harvest credentials.
- **Facebook (12.8%):** A significant channel for multi-generational social engineering, where synthetic media is often injected into private groups or promoted via fraudulent advertisements.
- **TikTok (12.5%):** The platform's algorithmic recommendation engine can inadvertently push synthetic content to highly specific demographics. Short-form deepfake videos often go viral before manual moderation teams can intervene.
- **YouTube (8.3%):** Used primarily for hosting longer-form deceptive content, such as fraudulent investment webinars or deepfaked expert interviews, which are then linked from other social platforms.

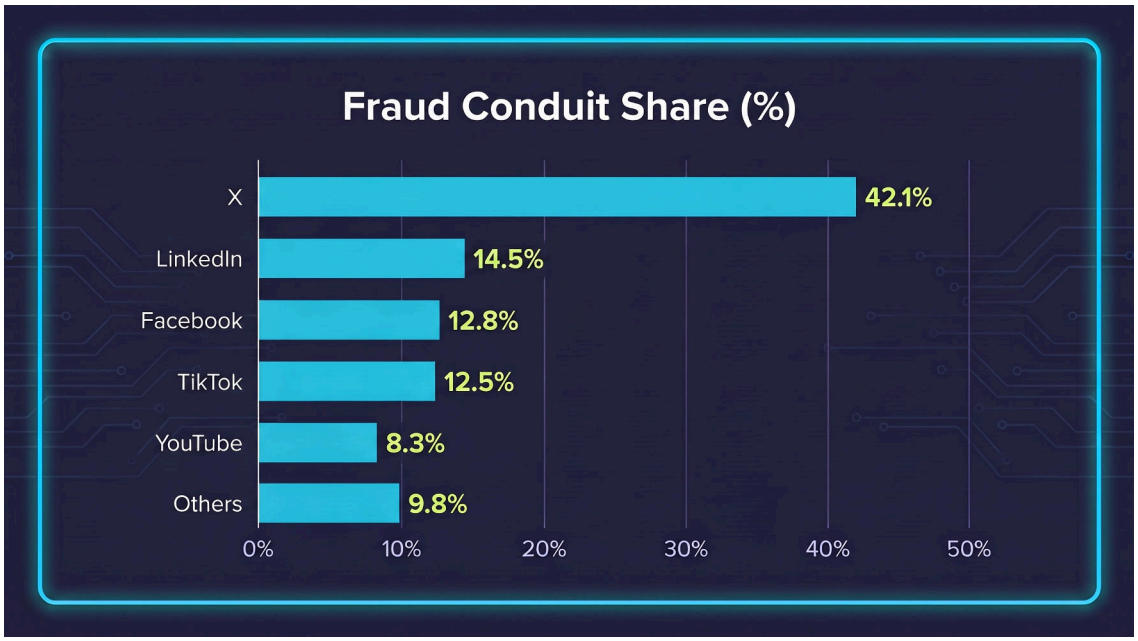


Figure 4: Distribution of deepfake fraud incidents across major social media platforms. - AI generated image

The Regulatory Shift to Platform Accountability The Regulatory Shift to Platform Accountability Social media platforms are no longer treated as passive intermediaries; global regulators are imposing affirmative duties of care and strict platform liability. Concurrently, European platforms face heavy compliance burdens under the Digital Services Act (DSA) and Article 50 of the EU AI Act, which mandate machine-readable provenance tracking and rapid response frameworks. As detailed in Section 7, platforms that fail to deploy automated deepfake detection capabilities risk direct regulatory liability and loss of legal safe-harbor protections. Adjacently, regulations like the TAKE IT DOWN Act target non-consensual intimate imagery (NCII), establishing clear platform liabilities for synthetic media in that domain.

6 Victim Profile & Sectors

Based on our observations spanning from January 2020 to August 2026, threat actors continue to expand their targeting aperture, shifting from opportunistic impersonation to highly targeted financial exploitation. Across all analyzed incidents globally, the top target sectors include **Private Individuals** (18.2%), **Corporate Entities** (15.8%), **Public Figures** (6.2%), and **Business Executives** (4.0%). While identifying exact victims is often challenging due to disclosure limitations, our data indicates a significant focus on high-fidelity impersonation of decision-makers.

However, a focused analysis of incidents specifically categorized as fraud reveals a stark escalation in the targeting of high-value corporate operations. Within these specifically categorized fraud incidents, targeting of the Corporate Entities sector accounts for 15.8%, underscoring a severe risk to organizational continuity. Furthermore, the Business Executive demographic constitutes 4.0% of the focused set. Figure 5 depicts the ranking of victim profiles and sectors within the fraud-specific incident set.

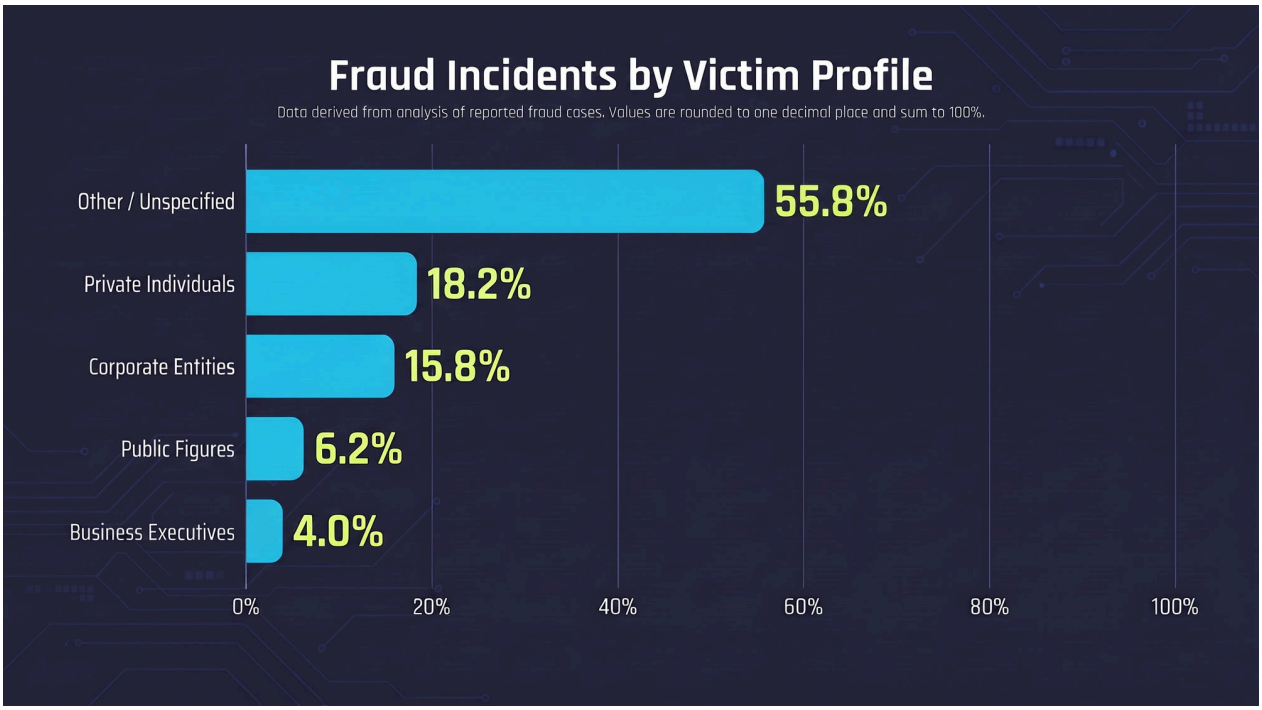


Figure 5: Ranking of victim profiles and sectors within the fraud-specific incident set. - AI generated image

The fastest-growing victim demographic consists of financial and enterprise targets—particularly corporate treasury and finance functions, banking, crypto and fintech platforms, and insurance providers. This growth is driven almost entirely by the massive payouts associated with corporate impersonation.

Key Finding

Threat actors are distinctly bifurcating their fraud operations:

- **Enterprise Wire-Transfer Targets:** Corporate decision-makers and high-net-worth individuals are targeted for next-generation Business Email Compromise (BEC), leveraging real-time **Video Synthesis** (42.7% of fraud techniques) and **Audio-based Deception** (21.9%) to authorize corporate treasury transfers.
- **Retail Investment-Scam Victims:** The Private Individual / General Public demographic (18.2% of fraud targets) is targeted for mass-scale crypto and investment scams.

While the enterprise tier focuses on wire-transfer targets, the retail tier relies on volume. Retail investment-scam victims are often targeted through social media distribution networks, capitalizing on identity fraud (17.6%) to promote deceptive cryptocurrency schemes. Recent 2026 law-enforcement advisories highlight mass-scale campaigns impersonating global leaders to market guaranteed-return financial products.

To counter the rise of synthetic identity fraud, global financial institutions are aggressively deploying liveness-detection and voice-biometric countermeasures. Nevertheless, as generative AI techniques evolve, defensive strategies must shift toward rigorous workflow verification to ensure that deepfake impersonations cannot unilaterally execute financial transactions.

7 Regulatory Landscape and Implications

The regulatory response to deepfake-enabled fraud is evolving from general AI, privacy and consumer-protection rules toward more explicit requirements for synthetic-media transparency, impersonation, content provenance, platform accountability and digital identity protection. There is no single global framework governing deepfake fraud; instead, organizations must navigate overlapping AI, financial-crime, privacy, consumer-protection and online-safety regimes. International bodies including Interpol, Europol, UNODC, the FBI/IC3 and the World Economic Forum Cybercrime Atlas increasingly support cross-border coordination and the development of common fraud typologies. These mechanisms are not themselves equivalent to legislation, but they contribute to emerging expectations around information sharing, identity protection and fraud prevention.

7.1 Regulatory Landscape

The regulatory direction is developing around five principal areas: transparency of synthetic content, liability for impersonation and deception, platform responsibilities, digital identity assurance, and stronger financial-sector authentication and monitoring controls. This increasingly connects deepfake detection with existing compliance frameworks rather than treating it as a standalone technology requirement.

In the European Union, the EU AI Act (Regulation 2024/1689) represents the most explicit regulatory framework for synthetic content. Article 50 transparency obligations apply from 2 August 2026, requiring relevant AI providers to support machine-readable marking of AI-generated content and establishing disclosure and labelling obligations for certain deepfakes and AI-generated content on matters of public interest. Note that the AI Omnibus amendment provides a targeted grace period until 2 December 2026 for the marking and detection obligations under Article 50(2) specifically, where the generative AI system was already on the market before 2 August 2026 – so the compliance timeline is not a single hard cutoff. Non-compliance can attract fines of up to €15 million or 3% of worldwide annual turnover, whichever is higher. These obligations sit within a wider EU legal web:

- **GDPR (Arts. 6 & 9):** Captures the biometric and personal data processing layer, strictly penalizing unauthorized biometric scraping or deepfake likeness generation lacking explicit data-subject consent.
- **Digital Services Act (DSA):** Imposes systemic risk mitigation and rapid notice-and-action obligations on major platforms hosting deceptive deepfakes.
- **eIDAS 2.0 & Financial Sector Rules:** Underpins the EU Digital Identity Wallet, establishing verification standards alongside EBA Guidelines and PSD2/PSD3 Strong Customer Authentication (SCA) to safeguard financial institutions against synthetic identity spoofing during remote onboarding.

Key Member States actively enforce localized national measures alongside EU frameworks:

- **Italy (National AI Law No. 132/2025 & Penal Code Amendments):** Italy became the first EU Member State to enact a comprehensive national AI framework through Law No. 132/2025. The statute introduced Article 612-quater into the Italian Penal Code, establishing a specific criminal offense for the "illicit dissemination of content generated or altered using artificial intelligence" without consent that is capable of deceiving as to its authenticity and causing unjust harm, punishable by 1 to 5 years imprisonment. Additionally, the statute amended Article 61 of the Penal Code to classify the malicious use of AI as a general aggravating factor in committing crimes, including market manipulation, fraud, and extortion.
- **France & Germany (National Electoral & Identity Enforcement):** France enforces strict criminal liability under its SREN Law (Digital Space Protection and Regulation Act), requiring platforms to deploy fast-track takedowns for non-consensual deepfakes and political deception. Germany relies on its Digital Services Enforcement Act alongside financial sector mandates (BaFin) requiring strict anti-spoofing countermeasures for remote video identification.

In the United Kingdom, there is no direct equivalent to the AI Act's deepfake-labelling regime. Instead, obligations are distributed across the Online Safety Act (platform duties around illegal and harmful content), the Fraud Act, and sector expectations set by the FCA and ICO. This creates a lighter-touch but still active compliance surface for UK-facing platforms

and financial institutions, and is a point of divergence from the EU worth monitoring as UK AI-governance proposals develop.

In the United States, regulation remains primarily sectoral at the federal level, complemented by aggressive state-level enforcement led by California:

- **FTC Impersonation Rule (16 CFR Part 461):** Provides direct federal enforcement tools against AI-enabled impersonation of governments and businesses, making it directly relevant to corporate Business Email Compromise (BEC) and social engineering; the FTC has separately signaled interest in extending impersonation protections to cover individuals, though this remains at the enforcement-posture stage rather than a settled statutory rule.
- **TAKE IT DOWN Act:** Notice-and-removal requirements for covered platforms became enforceable on 19 May 2026, requiring the removal of non-consensual intimate imagery—including synthetic or deepfake material—within 48 hours of a valid request. Violations are treated as FTC Act violations, carrying civil penalties of roughly $53,000 per violation. The FTC has actively begun enforcement, issuing formal warning letters to major platforms. While not primarily a financial-fraud statute, it establishes concrete federal platform accountability for hosting harmful synthetic media.
- **California AI Legislative Cluster:** California leads state-level enforcement through a comprehensive suite of targeted deepfake and synthetic media statutes:
  - **California AI Transparency Act (SB 942):** Mandates that covered generative AI providers embed machine-readable latent disclosures (provenance metadata) and offer user-facing manifest labeling tools for AI-generated video, audio, and images, backed by civil penalties of $5,000 per violation per day.
  - **Digital Replica & Likeness Protection (AB 1836 / AB 2602):** Establishes strict statutory civil liability and damages for the unauthorized creation, distribution, or commercial exploitation of an individual's digital voice or visual likeness (including post-mortem protections for deceased performers).
  - **Platform & Election Accountability (AB 2655 / AB 2355):** Imposes affirmative duties on major online platforms to detect, label, or block materially deceptive AI-manipulated political and financial deepfakes, establishing strict short-window takedown workflows.
- **Financial Sector Verification Standards:** FFIEC authentication guidance continues to shape operational expectations for how US financial institutions verify customer identity and detect account-takeover attempts, specifically requiring anti-spoofing controls against synthetic voice and video injection attacks.

In APAC, regulatory approaches vary significantly, blending prescriptive content-labeling mandates with strict criminal, electoral, and financial sector enforcement:

- **China:** Has established one of the world's most prescriptive regulatory frameworks through its Deep Synthesis Management Provisions and subsequent Measures for Labeling AI-Generated Synthetic Content (issued by the CAC and effective 1 September 2025). The measures impose mandatory dual-tier identification—requiring both explicit (visible/audible) disclaimers and implicit (embedded metadata) technical watermarks—across text, image, audio, and video generation providers.
- **India:** Enforces mandatory labeling and metadata retention for synthetically generated information through amendments to its IT Rules. On the financial sector side, the Reserve Bank of India (RBI) mandates anti-camera-injection and anti-spoofing controls for financial institutions executing remote Video-KYC onboarding to prevent synthetic identity fraud.
- **South Korea:** Maintains some of the region's most aggressive criminal and political prohibitions. Amendments to the Public Official Election Act strictly prohibit political campaign videos or advertisements created using generative deepfakes during the 90 days leading up to an election (subject to up to 7 years imprisonment). Furthermore, 2024 amendments to the Act on Special Cases Concerning the Punishment of Sexual Crimes escalated penalties for non-consensual deepfake generation and distribution to a maximum of 7 years in prison.
- **Singapore:** Addresses synthetic media threats through targeted statutory mechanisms. Under the Elections (Integrity of Online Advertising) Act, publishing, sharing, or boosting deepfakes depicting election candidates during election periods is strictly prohibited, subjecting non-compliant social media platforms to fines up to S$1

million. Additionally, Singapore's POFMA and Protection from Harassment Act (POHA) empower authorities to issue fast-track correction orders against synthetic financial deception and executive harassment.

- **Regional Trend:** Other key regional markets—including Japan, Australia, and Hong Kong (via HKMA operational directives on high-value transfers)—continue to expand AI-governance, online-safety, and banking authentication standards, establishing strict operational duties for enterprise verification architectures.

Latin America remains comparatively fragmented, though key regional markets are aggressively transitioning from general data protection rules toward dedicated AI governance, strict electoral prohibitions, and platform compliance obligations:

- **Brazil:** Brazil is developing a multi-layered framework for synthetic media through electoral regulation, data protection and emerging AI legislation. TSE Resolution No. 23.732/2024 specifically restricts the use of AI-generated or digitally manipulated audio and video ("deepfakes") in electoral propaganda and requires disclosure of AI-generated or manipulated content. Brazil's AI Bill 2338/2023 remains under legislative consideration and should therefore be treated as an emerging framework rather than current law. The LGPD also provides relevant protections where biometric or other personal data is processed in deepfake-related identity systems.
- **Mexico:** Addresses synthetic media threats through IP, data privacy, and emerging criminal frameworks. Recent Mexican Supreme Court rulings strictly limit intellectual property protections to natural human creators, while federal legislative amendments target non-consensual digital replicas, voice cloning, and AI-enabled financial impersonation under federal criminal and consumer protection codes.
- **Regional Trend:** Nations across the region—including Chile, Argentina, and Colombia—are advancing targeted legislative proposals addressing synthetic media fraud, image-rights protection, and platform provenance duties, signaling a region-wide shift toward concrete platform accountability.

7.2 Obligations for Stakeholders

The emerging framework creates different responsibilities across the ecosystem. AI providers and developers are increasingly expected to support content identification, transparency and appropriate safeguards. Platforms face growing expectations to label, restrict or remove certain harmful synthetic content and to maintain effective response mechanisms, with response-time requirements (e.g., TAKE IT DOWN's 48-hour window) becoming an emerging norm rather than the exception.

Financial institutions and enterprises face a more specific set of obligations. Beyond general fraud-prevention duties, this includes strong customer authentication requirements under regimes such as PSD2/PSD3 in the EU, FFIEC-aligned authentication expectations in the US, and typology guidance from FATF on AI-enabled financial crime. Where deepfakes are used to bypass identity verification, voice authentication or video-based KYC, institutions should expect regulators to treat inadequate controls against known, foreseeable synthetic-media risks as a compliance gap rather than a novel excuse. Biometric-specific rules – including the EU AI Act's restrictions on biometric categorization and existing regimes such as Illinois' BIPA – add a further layer of obligation where deepfake detection or identity verification tooling itself processes biometric data.

For organizations deploying deepfake detection, provenance and detection should be treated as complementary controls. Cryptographic provenance mechanisms such as C2PA can provide information on content origin and history, while independent detection remains relevant where provenance is absent, altered or stripped. This is consistent with the report's existing recommendation to combine provenance mechanisms with third-party AI detection rather than relying on provenance alone.

7.3 Enforcement & Liability

Enforcement is increasingly moving from voluntary principles toward measurable accountability, backed by concrete penalty regimes:

- **EU:** AI Act Article 50 breaches can attract fines of up to €15 million or 3% of worldwide annual turnover, whichever is higher, enforced by the European Commission and national authorities.

- **US:** The FTC is actively enforcing the TAKE IT DOWN Act's platform obligations (civil penalties of $53,000 per violation) and continues to use its Impersonation Rule against fraudulent impersonation schemes, with potential consumer redress and civil penalties.
- **China:** Regulators have moved from synthetic-media governance principles toward mandatory identification requirements with active implementation and enforcement.

Liability nevertheless remains distributed across the ecosystem and depends on jurisdiction and circumstances. Exposure may involve fraudsters, platforms, AI providers, financial institutions or enterprises where failures in prevention, authentication, disclosure or response contribute to harm.

7.4 Regulatory Implications

The overall trajectory is toward layered accountability rather than reliance on a single detection technology, with explicit penalty regimes now backing transparency and platform-accountability rules in the EU and US. Organizations operating internationally should prepare for increasing requirements around synthetic-content transparency, identity assurance (including digital identity frameworks such as eIDAS 2.0), provenance, KYC, transaction controls and incident response – while also tracking jurisdictions such as the UK and India, where obligations are emerging through existing statutes and sectoral guidance rather than a dedicated deepfake law. Deepfake detection is consequently best positioned as a supporting risk and compliance control within a broader verification architecture, rather than as a substitute for authentication, human review or established financial-crime controls.

Disclaimer: The legal and regulatory analysis contained in this section is provided for general informational and risk-assessment purposes only and does not constitute legal advice. Regulatory frameworks governing synthetic media are evolving rapidly; this summary is non-exhaustive and current as of the date of publication. Organizations should engage qualified legal counsel for binding compliance guidance in specific jurisdictions.

8 Real World Examples

Named High-Impact Cases & Quantified Losses

Our observations reveal a severe escalation in the sophistication and financial impact of synthetic media attacks between 2020 and 2026.

Threat actors have decisively shifted from basic email compromise to high-fidelity, real-time audiovisual impersonations of corporate leadership. This is exemplified by the January 2024 breach of Arup, a multinational engineering firm, resulting in a $25.6 million loss across fifteen transfers. In this attack, a finance worker was deceived by a video conference call where the corporate Chief Financial Officer and several senior colleagues were entirely impersonated using real-time generative video and voice deepfakes.

Similar tactical frameworks have been deployed globally. In mid-2024, attackers targeted a senior executive at Ferrari via WhatsApp, using a highly convincing neural voice clone of the luxury automotive brand's Italian Chief Executive Officer to request a confidential currency-hedge transaction[3]. The scam was narrowly thwarted when the targeted executive asked a verification question about a recently recommended book. Other incidents include an unsuccessful synthetic video call scam targeting the CEO of WPP[4], which was intercepted by security controls, and a high-yield investment scam in Singapore utilizing deepfakes of government officials[5], resulting in a S$4.9 million loss for a single victim.

Beyond executive impersonation, our observations highlight a sustained assault on banking Know Your Customer (KYC) protocols. Threat actors routinely inject synthetic faces directly into verification API streams to secure fraudulent loans, classifying 17.6% of all documented fraud deepfakes specifically as identity fraud. Additionally, **Audio-based Deception** (21.9%, encompassing Audio Manipulation and Voice Clones) is heavily utilized in vishing campaigns targeting corporate IT helpdesks to bypass Multi-Factor Authentication (MFA).

The financial damage of these campaigns is severe. Our observations confirm that the incidents analyzed in this specific vertical were financially motivated fraud, with **Business Executives** comprising 4.0% of targeted entities, while broader **Corporate Entities** accounted for 15.8%.

End-to-End Attack Chain Breakdown

Forensic analysis of these incidents exposes a highly systematized, universal four-stage attack lifecycle:

1. **Target Selection & Reconnaissance:** Threat actors meticulously harvest OSINT to build synthetic profiles. According to our observations, the raw material for these attacks is primarily sourced from X, LinkedIn, and other major social platforms. Attackers extract high-quality audio and video of C-suite executives from investor calls, podcasts, and corporate interviews to train their models.
2. **Media & Vector Synthesis:** Attackers generate the synthetic payload. Based on identified media modalities, **Video Synthesis** represents the vast majority of our observed attacks (42.7%, comprising Video Manipulation and GenAI Video), followed by **Unspecified or Other** techniques (24.0%) and **Audio-based Deception** (21.9%). Threat actors utilize advanced diffusion models for video generation and neural frameworks for low-latency voice cloning, establishing virtual camera injection hooks to pipe synthetic media directly into enterprise conferencing software.
3. **Strategic Delivery:** Delivery mechanisms have pivoted away from traditional email. Threat actors now prefer direct vishing, WhatsApp audio notes, and spoofed video conferencing links. By exploiting trusted internal communication channels, the delivery method inherently lowers the target's suspicion threshold.
4. **Exploitation & Wire Authorization:** The final phase neutralizes standard financial controls. By presenting synthetic C-suite authorization in real-time, attackers successfully bypass dual-authorization and verbal verification protocols, forcing the immediate execution of unauthorized wire transfers.

[3] Source: https://www.drive.com.au/news/ferrari-ceo-impersonated-ai-deepfake-scam/
[4] Source: https://www.theguardian.com/technology/article/2024/may/10/ceo-wpp-deepfake-scam
[5] Source: https://www.straitstimes.com/singapore/victim-loses-at-least-4-9m-in-scam-involving-deepfakes-of-pm-wong-government-officials

Key Finding

Our observations demonstrate that text-based fraud is rapidly being replaced by synthetic audiovisual attacks. With **Video Synthesis** (42.7%) and **Audio-based Deception** (21.9%) comprising over 64% of deepfake fraud vectors, traditional employee security awareness training–which relies on spotting typographical errors–is structurally insufficient.

The case files below were curated as high-signal, representative examples of deepfake-enabled report – each illustrating the tradecraft, targeting and financial impact analysed in this dossier.

The following real-world case studies illustrate the accelerating sophistication of AI-driven fraud across the digital threat landscape. Selected by analysts at identifAI, these incidents demonstrate a strategic shift from technical hacking to the manipulation of human trust and identity verification systems. They underscore how rapidly synthetic media is being weaponized to execute high-impact financial and credential-based scams.

Indonesian Financial Institution: 1,100 Deepfake Fraud Attempts

Figure 6: Visualizing the scale of synthetic identity injection in mobile banking platforms. - AI generated image

In mid-2024, an Indonesian financial institution experienced over 1,100 attempts of deepfake fraud within its mobile app (*Source: Frontier Enterprise, "Deepfake Fraud: AI's Impact on Financial Institutions"* - https://www.frontier-enterprise.com/deepfake-fraud-ais-impact-on-financial-institutions/). Attackers used manipulated IDs and synthetic media to bypass biometric verification protocols. While national macroeconomic loss models have projected Indonesian deepfake fraud impacts at US $138.5 million, this institution successfully mitigated the scale of direct realized losses through its security response. This incident directly validates the **End-to-End Attack Chain** analysis provided in this report, specifically the "Media & Vector Synthesis" stage where high-volume synthetic injection hooks are used to bypass automated KYC onboarding at scale.

Arup: $25.6 Million Video Conference Breach

Figure 7: Conceptual visualization of a deepfaked video conference used to authorize a fraudulent $25.6 million transfer. - AI generated image

In one of the most prominent cases of 2024, an employee at Arup was tricked into transferring $25.6 million during a deepfaked video conference (*Source: CNN, "Deepfake CFO scam: Hong Kong police say clerk was tricked into paying $25 million"* - https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk). Every individual on the call, including the person who appeared to be the company's Chief Financial Officer, was a computer-generated impostor.

This incident serves as a definitive case study for the **Exploitation & Wire Authorization** phase described in Section 3, illustrating how synthetic C-suite presence can be used to override internal financial controls. It further underscores the findings in the **Victim Profile** analysis (Section 5), where high-value corporate targets are increasingly subjected to multimodal attacks that bypass traditional verbal and visual verification protocols.

1 in Every 100 Identity Check Failures Involves Deepfake Media

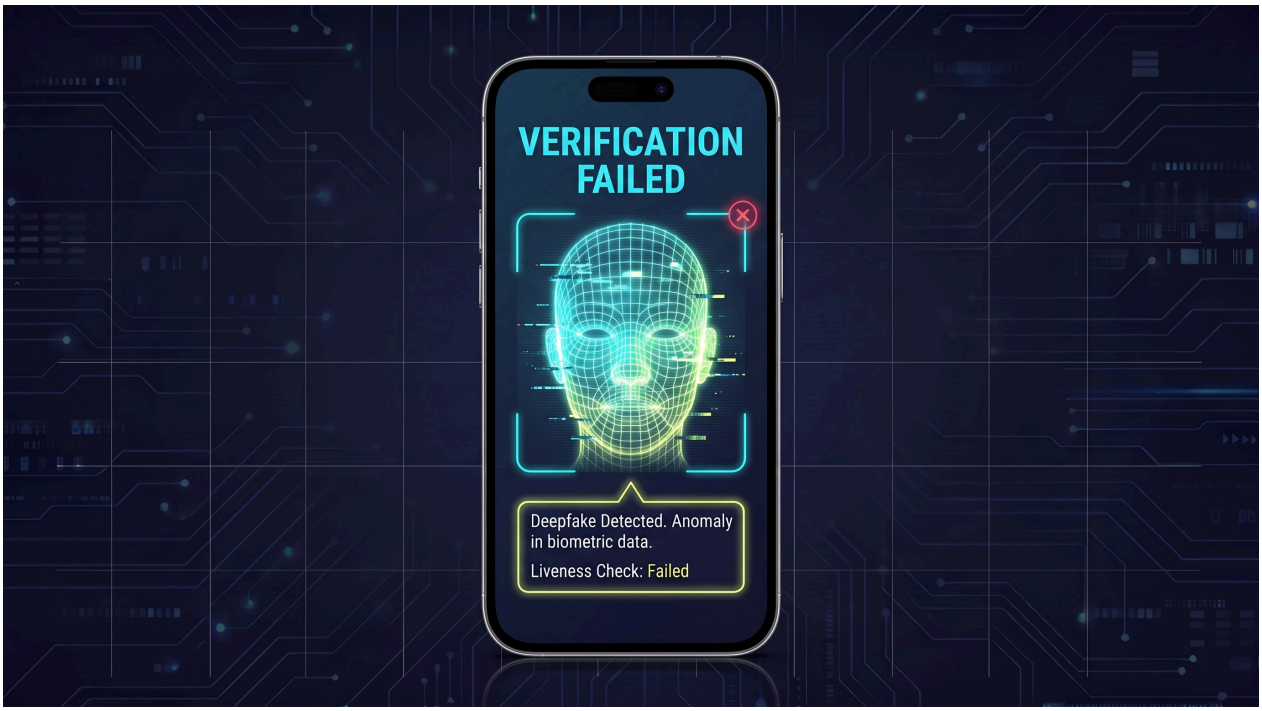
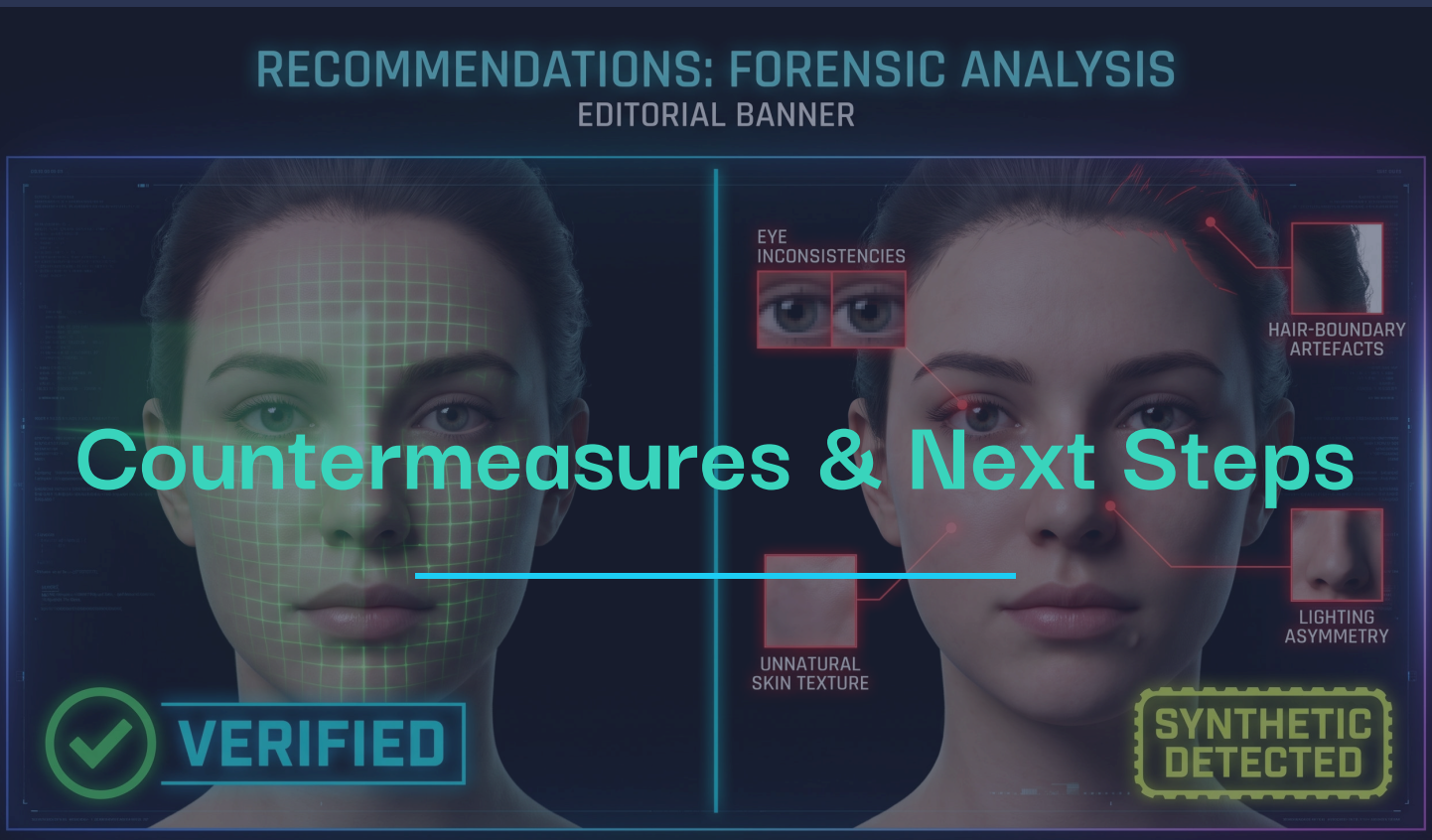


Figure 8: Visualization of a failed biometric identity check due to synthetic media injection. - AI generated image

Historically relied upon as definitive proof of human presence, identity verification protocols are being thoroughly compromised by realistic AI generation. Recent data indicates that one in every 100 identity check failures now involves a deepfake document, image, or liveness video (*Source: Identity Week, "1 in every 100 identity check failures involves deepfake media" - https://identityweek.net/1-in-every-100-identity-check-failures-involves-deepfake-media/*).

This represents a 180% rise in deepfake-assisted identity fraud attempts as attackers scale their use of synthetic media to bypass automated KYC onboarding. This trend directly correlates with the **Attack Taxonomy** (Section 3), where 'Identity Fraud' (17.6%) is used as a gateway for broader financial exploitation. As noted in the **Victim Profile** analysis (Section 5), these failures underscore the inadequacy of standalone biometric checks in the face of industrial-scale synthetic generation.

Across these curated incidents, a distinct pattern emerges: cybercriminals are systematically pivoting from traditional, code-based exploits toward manipulating human trust and biometric verification systems via synthetic media. Whether bypassing dozens of automated identity checks or convincing users to authorize devastating financial transfers, the weaponization of AI is maximizing the scale and success rate of modern fraud. As identifAI analysts note, mitigating this escalating threat requires security paradigms that no longer rely solely on visual proof, but instead integrate robust, multi-layered deepfake detection capabilities.



Countermeasures & Next Steps

9 Strategic Countermeasures

Key Finding

The convergence of **Video Synthesis** and **Image Synthesis** is accelerating **Identity Fraud** during synthetic onboarding bypasses. All stakeholder segments must actively align their deepfake fraud taxonomies with the latest Interpol and Europol advisories to maintain a unified defense.

Banks, Payment & KYC Providers

Recent data highlights a shift toward industrialized, high-volume synthetic fraud. The threat has rapidly evolved from early executive impersonations to massive, coordinated operations, such as the December 2025 dismantling of a transnational cryptocurrency scam that used deepfakes to defraud victims of more than €700 million[6]. Additionally, the June 2026 breach of Indian Non-Banking Financial Companies (NBFCs)[7]–where synthetic media was used to successfully bypass Video KYC protocols and resulted in losses of 20 Indian Rupees Crore ($2.4M USD)–underscores the critical vulnerability of remote onboarding systems. Across all these cases, traditional identity verification perimeters are increasingly breached by sophisticated virtual injection techniques.

1. **Deploy camera injection defenses:** Enforce multi-modal biometric liveness detection thresholds to identify and block virtual camera loops and synthetic face overlays. Crucially, because client-side hardware and mobile applications must be assumed compromised by injection attacks, organizations must implement strict backend countermeasures. This requires routing cryptographically signed sensor payloads to the backend for server-side evaluation, utilizing multimodal models specifically tuned to detect underlying generative artifacts rather than relying on standard biometric matching. Institutions must prioritize these secure, active liveness checks to counter Identity Fraud, which comprises 17.6% of observed fraud objectives.
   Time-horizon: Immediate.
   Indicator: Measurable reduction in synthetic onboarding bypasses.
2. **Implement transaction-anomaly controls:** Integrate call-back and out-of-band verification protocols tailored for high-risk transfers. To counter advanced impersonation, payment providers must deploy behavioral biometrics alongside dedicated deepfake detection–relying on multimodal models to identify underlying generative anomalies–to flag anomalous authorized push payments before execution, establishing a true zero-trust environment for corporate withdrawals.
   Time-horizon: 12-Month.
   Indicator: A sustained, measurable reduction in unauthorized push payment fraud percentages.

Enterprise Finance & Risk Teams

Threat actors are heavily harvesting targeting material from public websites and dataleak to orchestrate Business Email Compromise (BEC). Our observations indicate that attacks targeting **Corporate Entities** and **Business Executives** frequently utilize **Audio-based Deception** (21.9%), as seen in recent vishing attempts against high-profile technology executives.

1. **Institute specialized BEC and vishing playbooks:** Develop targeted incident response protocols tailored to synthetic media threats. Organizations must mandate out-of-band verification tooling for high-value financial transactions alongside rigorous staff training focused on identifying voice cloning artifacts.
   Time-horizon: 12-Month.
   Indicator: 100% staff completion of AI voice anomaly awareness training.

[6] Source: https://www.europol.europa.eu/media-press/newsroom/news/international-takedown-of-cryptocurrency-fraud-network-laundering-over-eur-700-million

[7] Source: https://the420.in/deepfake-video-kyc-banking-fraud-ai-identity-verification-threat/

2. **Enforce dual-authorisation for wire transfers:** Traditional multi-party sign-off and out-of-band verification are no longer sufficient on their own. Risk teams must fully decouple final wire authorizations from initial communication channels to isolate impersonation attempts. Crucially, this decoupling must be reinforced through seamless API integrations with advanced deepfake detection platforms. By embedding multimodal models that analyze real-time communications for underlying generative artifacts—rather than relying on legacy biometric matching—organizations can programmatically intercept synthetic voice and video injections, establishing an automated, zero-trust perimeter around high-value financial transactions.
   Time-horizon: Immediate.
   Indicator: Zero successful bypasses of enterprise financial protocols.

Policymakers & Regulators

To counteract the widespread distribution of synthetic media, regulatory frameworks must evolve beyond voluntary guidelines. With X (42.1%) remaining a prominent social network for threat reconnaissance, stringent platform accountability and verifiable media pipelines are strictly required.

1. **Standardize provenance mechanisms:** Mandate the integration of cryptographic standards such as C2PA across all hardware and software layers. Regulators must enforce media authenticity at the point of capture to establish a baseline of data provenance and disrupt the proliferation of synthetic identities at scale. However, while cryptographic provenance is a critical foundational step, it is inherently insufficient on its own. Threat actors can easily bypass implicit or explicit watermarking through metadata stripping, format conversion, or by exploiting the 'analog hole' (such as screen recording). To guarantee comprehensive coverage across the full spectrum of synthetic media threats, these standards must be paired with independent deepfake detection models. Relying on multimodal architectures to analyze underlying generative artifacts ensures a robust, active layer of defense even when cryptographic seals are compromised or entirely absent.
   Time-horizon: Strategic.
   Indicator: Broad adoption of C2PA standards across consumer digital supply chains and most of all integration to third party AI detectors.
2. **Finalize strict platform liability frameworks:** Regulators must forcefully implement authorized-push-payment (APP) liability frameworks to ensure strict accountability across both financial and digital platforms. It is no longer acceptable for institutions to pass the cost of synthetic media fraud onto the end victim. By legally shifting the financial burden onto platforms that fail to deploy advanced countermeasures—such as multimodal artifact detection—regulators can mandate a higher standard of care. Concurrently, regulators must actively foster international collaboration and mandate cross-border fraud intelligence sharing, specifically regarding emerging deepfake evasion techniques, to successfully track and dismantle distributed threat actors on a global scale.
   Time-horizon: Strategic.
   Indicator: Implementation of synchronized, cross-border incident registers.

10 Conclusions

Key Finding

While the theoretical applications of synthetic media and advanced digital deception span a broad spectrum of malicious intent—ranging from state-sponsored market manipulation and corporate espionage to hacktivism and geopolitical sabotage—our current telemetry indicates a markedly narrower operational focus. Empirical data from the reporting period reveals that the threat landscape is almost exclusively dominated by campaigns driven by direct financial gain, heavily leveraging **Video Synthesis** (42.7%) and **Identity Fraud** (17.6%) to bypass traditional verification controls and deceive institutional targets. Threat actors are consistently prioritizing immediate monetization, deploying high-fidelity voice cloning and face-swapping technologies primarily to bypass biometric authentication, orchestrate sophisticated Business Email Compromise (BEC) schemes, and execute unauthorized fund transfers. Consequently, although the potential for ideological or disruptive attacks remains a critical future risk, the present reality of these advanced threat vectors is defined by a singular, pragmatic pursuit of illicit financial extraction.

Synthesis of Critical Findings

1. **Recalibrate** liveness detection protocols to counter advanced visual and audio synthesis. [Impact Horizon: Immediate]. Our observations indicate **Video Synthesis** (42.7%) and **Audio-based Deception** (21.9%) drive modern business email compromise. The December 2025 dismantling of a transnational cryptocurrency scam that leveraged deepfakes to defraud victims of more than €700 million[8] highlights this critical vulnerability. *Measurable Strategic Outcome: Achieve full deployment of multi-modal liveness verification across enterprise transaction flows.*
2. **Neutralize** malicious content across dominant social platforms. [Impact Horizon: 12-Month]. Fraud incidents in our observations are predominantly discovered via X (42.1%) and LinkedIn (14.5%). Threat actors exploit these platforms to scrape corporate hierarchy data and distribute synthetic identity lures targeting **Corporate Entities** (15.8%). *Measurable Strategic Outcome: Decrease successful synthetic identity onboarding rates and fraudulent social-media driven traffic by strengthening automated perimeter defenses.*

Strategic Business Implications

1. **Quantify** synthetic fraud risk within enterprise risk frameworks to mitigate severe financial and reputational damage. [Impact Horizon: Immediate]. With the United States accounting for 34.09% of incidents and global operations making up 27.0%, cross-border financial institutions face unprecedented exposure. Law-enforcement advisories from Europol and Interpol warn of escalating losses from biometric KYC bypass attacks. *Measurable Strategic Outcome: Integrate synthetic threat modeling into quarterly financial risk assessments to prevent catastrophic wire-transfer fraud.*
2. **Overhaul** executive protection and brand management strategies. [Impact Horizon: Strategic]. **Corporate Entities** (15.8%) and **Business Executives** (4.0%) face sophisticated targeting in our observations. Campaigns frequently utilize **Audio-based Deception** (21.9% of fraud types) of senior corporate leaders to authorize illicit payments or manipulate stock prices. *Measurable Strategic Outcome: Establish continuous, cryptographic biometric baselines for all C-suite personnel to enable real-time deepfake audio and video rejection.*

[8] Source: https://www.europol.europa.eu/media-press/newsroom/news/international-takedown-of-cryptocurrency-fraud-network-laundering-over-eur-700-million

Forward-Looking Outlook & Regulatory Alignment

1. **Enforce** strict compliance with emerging synthetic media mandates. [Impact Horizon: 12-Month]. Article 50 of the EU AI Act and the US FTC's 2024 Impersonation Rule mandate transparency and penalise AI-enabled fraud. Enterprises must audit their verification stacks to ensure alignment with these jurisdictional obligations. Adjacently, legislation like the TAKE IT DOWN Act targets non-consensual intimate imagery, focusing platform liability on synthetic media in that specific domain. *Measurable Strategic Outcome: Complete comprehensive regulatory gap assessments for deepfake transparency and fraud prevention standards across all operational geographies.*
2. **Adopt** cryptographic provenance frameworks to secure digital supply chains. [Impact Horizon: Strategic]. As **Video Synthesis** (42.7%) and **Image Synthesis** (8.5%) continue to mature, adopting standards like C2PA becomes critical for ecosystem resilience. Without verifiable digital provenance, distinguishing authentic corporate communications from synthetic identity fraud (17.6%) will become untenable. *Measurable Strategic Outcome: Implement C2PA-compliant watermarking and metadata tracking across official external corporate communications.*

Ultimately, the proliferation of both open-source repositories and low-cost consumer applications has fundamentally democratized access to deepfake technologies. By removing the traditional barriers of technical expertise and expensive hardware, this ecosystem has enlarged the magnitude of synthetic media production to an unprecedented scale. We are no longer facing isolated incidents of digital forgery, but rather a massive, continuous influx of fabricated content.

As a direct consequence of this volume and quality, traditional defense mechanisms are failing. Historically, the burden of identifying manipulated media fell to trained deepfake analysts who could meticulously spot subtle artifacts, unnatural lighting, or audio inconsistencies. Today, however, generative models have become so sophisticated and refined that these telltale signs have largely been eliminated. The technology has crossed a threshold where human experts, regardless of their training or experience, are increasingly unable to reliably distinguish synthetic illusions from baseline reality.

Given this immense scale and the incredibly advanced nature of modern generative AI, relying on manual human review is no longer a viable safeguard. To effectively combat this evolving threat landscape, we must fight AI with AI. There is an urgent, non-negotiable need to develop, standardize, and deploy robust, automated detection technologies capable of accurately identifying and flagging deepfakes at massive scale.

11 Methodology

The threat intelligence presented in this report is synthesized through identifAI's multi-layered observation framework, which continuously monitors global English-language news media, journalistic repositories, and public social media platforms. The reporting window spans from January 2020 through August 2026.

Incident Harvesting and Data Pipeline Our automated harvesting architecture aggregates raw digital telemetry from open-source intelligence (OSINT) channels, verified media feeds, and social network public streams. An incident is strictly defined as an observed case of a deepfake being utilized in the real world, as documented by global English-based news outlets or identified within public social media posts.

LLM-Powered Classification and Taxonomy Curation Harvested telemetry undergoes advanced Large Language Model (LLM) classification pipelines to evaluate contextual relevance, strip duplicate entries, and categorize events across technical carriers (video, audio, still image, identity fraud), target sectors, and distribution platforms. Extracted data points are cross-referenced against standardized threat taxonomies, and statistical normalization methods convert raw empirical counts into comparative percentage distributions.