International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

AI IMPERSONATION PROBLEM: A REVIEW ON DEEPFAKES, SYNTHETIC IDENTITY, AND THE COLLAPSE OF DIGITAL TRUST

V. Snitha Sri[1] | M. Sri Srushti[2] | N. Sharvani [3] | N. Sathvika [4]

Department of Bachelors of Computer Science, Vishwa Vishwani Institute of Systems and Management, Hyderabad, India

ABSTRACT

The rapid democratization of generative artificial intelligence and high-fidelity neural synthesis has eroded the traditional premise that biometric audiovisual recordings serve as indisputable proof of reality. Deepfakes—spanning real-time voice cloning and neural facial reenactment—have shifted the cyber threat landscape from exploiting software vulnerabilities to compromising human perceptual trust. This study investigates the socioeconomic, psychological, and institutional ramifications of synthetic identity generation and the subsequent collapse of digital trust. Utilizing a descriptive research methodology based on multi-source secondary data, this paper explores the anatomy of AI impersonation, with particular attention to social engineering in the financial sector, legal ambiguities surrounding victim liability, and the profound trauma inflicted upon targeted individuals and families. The findings indicate an asymmetric arms race between generative synthesis and post-hoc forensic detectors, emphasizing that isolated algorithmic filtering is insufficient. The study concludes that mitigating digital identity decay requires a defense-in-depth model that combines cryptographic provenance (such as C2PA standards), multi-modal liveness verification, dynamic institutional protocols, and statutory reforms clarifying victim liability.

Keywords: *Deepfakes, Synthetic Identity, Voice Cloning, Digital Trust, Social Engineering, Cryptographic Provenance.*

1.INTRODUCTION

In the contemporary digital economy, personal identity verification and institutional communication depend heavily on digital representations: video feeds, recorded voice notes, and photographic credentials. Historically, photographic and audio recordings served as reliable evidentiary records of physical presence and authentic human intent. However, the advent of deep learning architectures—specifically Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and neural diffusion models—has altered this paradigm fundamentally. When any individual's likeness and vocal timbre can be photorealistically synthesized with minimal source data, digital evidence forfeits its intrinsic presumption of authenticity.

The critical problem of AI impersonation does not lie primarily in software code exploitation or network firewall penetration; rather, it directly targets human cognitive heuristics and interpersonal trust. Cyber criminals weaponize real-time AI-generated video and acoustic voice clones to execute high-impact spear-phishing, manipulate executive directives, siphon institutional capital, and blackmail private citizens. Behavioral Anomaly Analytics: Shifting from static filtering to dynamic analysis—evaluating baseline communication patterns, request velocity, and unexpected authorization paths rather than media authenticity alone. Establishing structural safeguards against this systemic erosion of digital trust has become a vital priority for cybersecurity, legal policy, and sociotechnical research.

2. REVIEW OF LITERATURE

Scholarly exploration of synthetic media highlights the dual progression of generative tools and systemic societal exposure :

- **Paris & Donovan (2020)** analyzed the spectrum connecting crude manipulations ('cheapfakes') to high-grade deepfakes, noting that visual deception relies as much on social context and pre-existing cognitive biases as on algorithmic sophistication.
- **Lyu (2020) and Wang et al. (2023)** evaluated forensic detection frameworks, observing that deep convolutional networks frequently overfit to known artifacts (e.g., unnatural blinking patterns, boundary warping, and frequency spectrum anomalies). Their findings emphasize that generative models rapidly adapt to eliminate detectable flaws, creating an asymmetric arms race between generation and forensic classification.
- **Marchal et al. (2024)** classified operational vectors in generative AI misuse, identifying corporate executive impersonation and non-consensual biometric exploitation as the most prevalent and damaging real-world fraud tactics across global commercial sectors.

3. RESEARCH GAP

While extensive computer vision literature addresses post-hoc deepfake classification techniques, scholarly attention remains fragmented regarding the socioeconomic / human consequences of synthetic fraud. Research in present time isolates technical artifact detection without examining how rapid generative advances outpace existing liability jurisprudence, leave victims financially and psychologically vulnerable, and systematically disintegrate enterprise-level authentication frameworks. This paper bridges that gap by connecting the technical realities of basic synthesis with socio-legal analysis, practical experience of fraud trends, and architectural defenses.

4. OBJECTIVES OF THE STUDY

The data submitted to the National Assembly showed that since 2020, 639,564 complaints have been filed with FIA's cybercrime cell, which verified 414,260 of them. These resulted in 73,825 inquiries and 5,713 cases filed in courts.

> A breakdown of the data showed that of the 625 people arrested in 2020, 20 were convicted. The number was 38 out of As per the data, the number of complaints fell in 20 after rising for three consecutive years. However, according to the spokesperson of FIA's cybercrime wing, the data only reflected complaints in the first three quarters of 2024.

The number of complaints in the entire year stood at around 160,000.

The spokesperson said the number of complaints was still "quite less" compared to the number of internet users in the country.

The primary objectives directing this investigation are:

- To evaluate the mechanisms of AI impersonation, including neural voice synthesis and real-time facial reenactment.
- To do the legal ambiguities regarding fraud victim liability and the psychological burdens imposed on affected parties and families
- To examine how deepfake-driven social engineering disrupts institutional finance and separate privacy.
- To suggest multi-layered governance and technical remedies, focusing on cryptographic provenance and adaptive authentication.
- To improve empirical trends and statistical distributions of emerging deepfake threats across sectors.

5. RESEARCH METHODOLOGY

- Research Design: conceptual, and diagnostic research.
- Nature of Data: Secondary data drawn from academic journals, institutional threat reports, and regulatory advisories.
- Sources of Data: Peer-reviewed publications, global cybersecurity bulletins (e.g Deloitte), statutory policy frameworks, and regulatory notices.
- Analytical Method: Conceptual synthesis evaluating generative threat surfaces, structural socioeconomic impacts, and architectural mitigation models.

6. CONCEPTUAL BACKGROUND

Artificial Intelligence Impersonation: The deployment of synthetic generative algorithms to clone, simulate, or replace a specific individual's physical appearance, facial micro-expressions, or vocal characteristics with photorealistic accuracy.

Synthetic Identity: Fabricating entirely novel, composite, or cloned identities by blending real, stolen, and computationally generated biometric attributes to bypass automated onboarding, remote proctoring, and KYC security protocols.

Digital Trust: The quantifiable confidence placed by citizens, commercial organizations, and digital systems in the authenticity, integrity, and non-repudiation of digital communications, transactions, and electronic records.

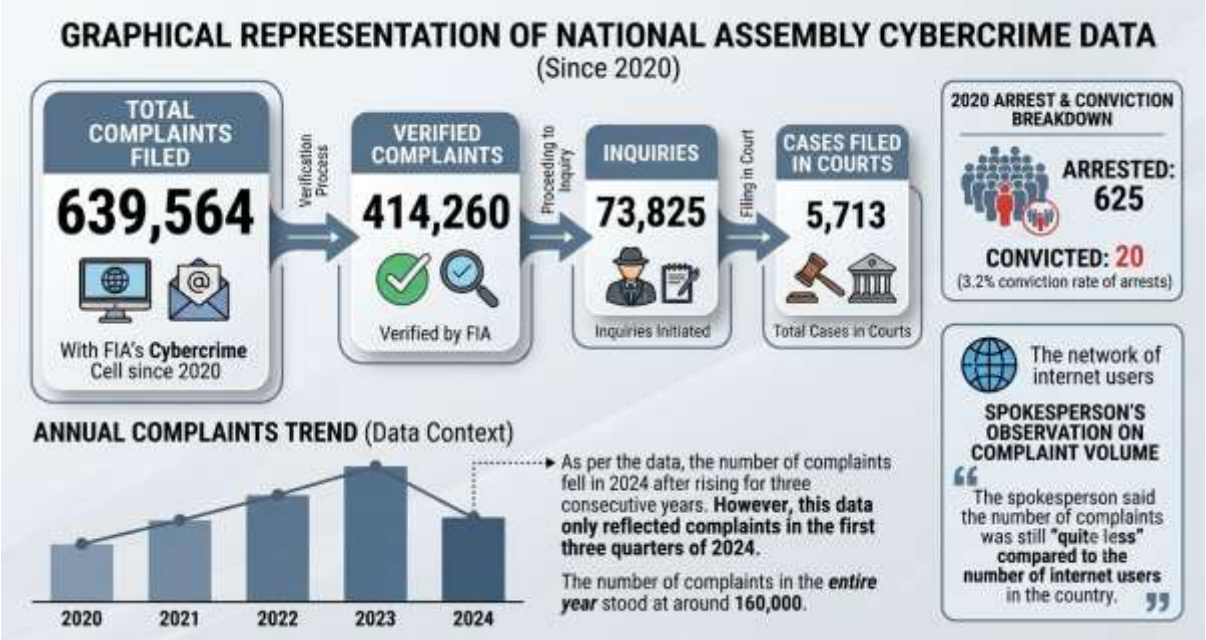
7. ANALYSIS AND DISCUSSION

7.1 Real-World Harm: How People and Organizations Face the Threat

The evolution from crude, asynchronous video manipulation to low-latency, real-time generative cloning has introduced severe threats across corporate, individual, and familial domains:

- **Virtual Kidnapping & Psychological Terror:** Scammers harvest seconds of audio from social media videos to clone the voices of young family members. They stage realistic calls to parents or grandparents claiming severe car accidents, police detention, or abductions, demanding immediate ransom. Beyond the direct financial loss, victims suffer enduring psychological trauma and post-traumatic stress.
- **Corporate Financial Fraud & Executive Impersonation:** Cybercriminals deploy multi-person synthetic video conferences and cloned vocal instructions to direct finance teams to execute urgent, off-ledger wire transfers. Notable enterprise attacks have resulted in losses exceeding $25 million in single events, where employees operated under the

conviction they were engaging directly with genuine chief financial officers.



- **The Victim Liability Void:** Present commercial banking and regulatory frameworks define unauthorized electronic fund transfers around stolen credentials. When an authorized account holder executes a transaction under the influence of a convincing audiovisual deepfake, financial institutions frequently claim gross negligence on the part of the account holder, denying reimbursement and leaving the victim fully liable for catastrophic financial losses.

7.2 Empirical Distribution and Growth Trends

Threat intelligence reports demonstrate an alarming upward trajectory in synthetic identity exploitation across global digital identity verification channels.

Table 1: Deepfake Fraud Trends & Financial Sector Impacts (2022–2024)

Metric / Threat Dimension	2022	2023	2024	Primary Impact Vector
Share of Video Biometric Fraud	~10%	21%	40%	Identity Verification / KYC Bypass
Businesses Encountering Deepfakes	29%	37%	49%	Social Engineering & Phishing
YoY Incident Volume Escalation	Baseline	+1,000%	+400%	Cumulative 40x Growth (Fintech focus)
Avg Loss per Affected Financial Entity	~$180k	~$390k	~$603k	Wire Fraud, Unauthorized Asset Transfers

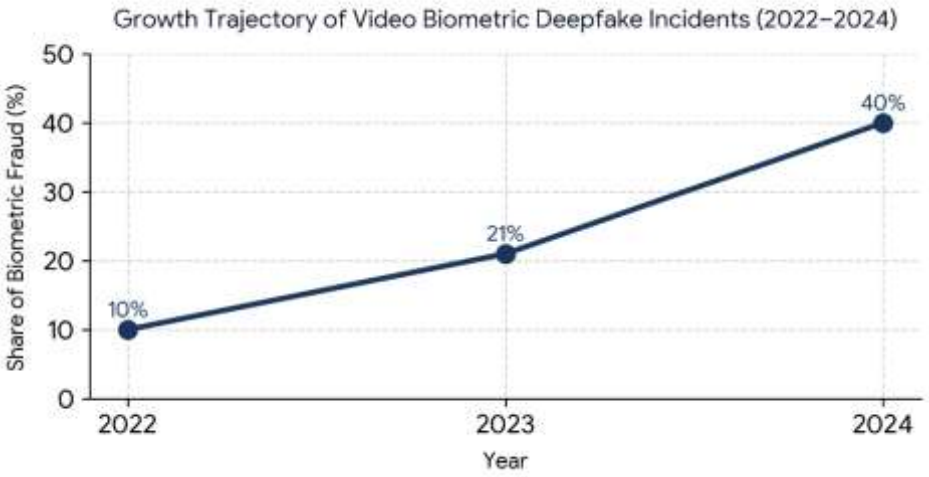


Figure 1: Growth Trajectory of Video Biometric Deepfake Incidents (%)

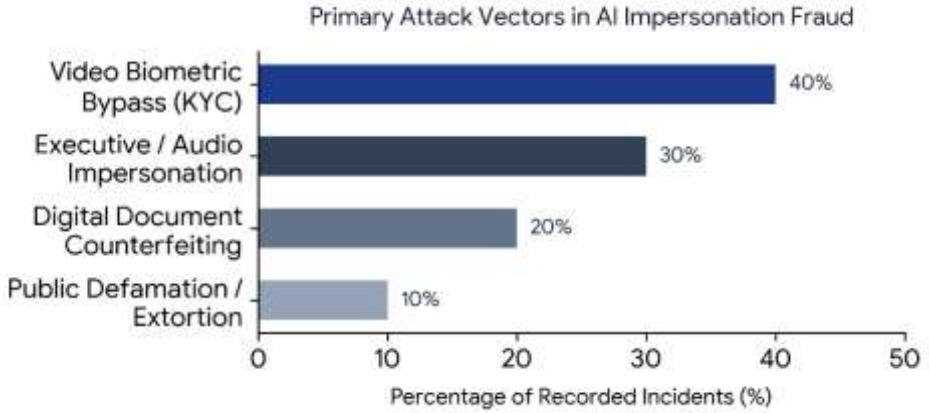


Figure 2: Distribution of Attack Vectors in AI Impersonation Fraud

> A breakdown of the data showed that of the 625 people arrested in 2020, 20 were convicted. The number was 38 out of as per the data, the number of complaints fell in 20 after rising for three consecutive years. However, according to the spokesperson of FIA's cybercrime wing, the data only reflected complaints in the first three quarters of 2024.

> The number of complaints in the entire year stood at around 160,000.

7.3 The Failure of Purely Algorithmic Detection

Organizations frequently rely on automated post-hoc classifiers trained to spot anomalies in media files. However, these systems exhibit structural fragility in real-world deployment:

1. Detection Evasion: Minor compression artifacts, slight Gaussian blurring, or deliberate adversarial perturbations degrade automated detection accuracies from over 90% in lab conditions to near chance levels in production environments.
2. Arms Race Asymmetry: As soon as forensic researchers publish a new metric for detecting synthetic artifacts (such as irregular eye blinking or corneal reflections),

generative AI developers incorporate those constraints directly into the loss function of their discriminator models, systematically eliminating the artifact in next-generation releases.

8. MAJOR FINDINGS

• Perceptual Weaponization: Deepfakes bypass traditional boundary firewalls by weaponizing authority, familiarity, and urgent human emotional reflexes.

• Structural Limits of Post-Hoc Filtering: Retrospective deepfake classifiers cannot sustain defensive integrity against continuous adversarial learning without hardware-anchored origin verification.

• Exponential Threat Escalation: Driven by accessible open-source neural toolkits, video biometric fraud expanded fortyfold from 2022 to 2024, becoming a dominant KYC bypass mechanism.

• Legal Redress Vacuum: Existing banking regulations and criminal codes lack precise statutory accommodation for generative deception, frequently imposing financial liability and deep psychological distress onto duped victims.

9. SUGGESTIONS AND RECOMMENDATIONS

• **Modernize Statutory Liability Frameworks:** Legislative and banking authorities must modernize electronic fund transfer regulations so that victims targeted by sophisticated neural impersonation are not automatically disqualified from reimbursement under obsolete negligence standards.

• **Adopt Cryptographic Provenance (C2PA):** Implement hardware-rooted cryptographic watermarking and Coalition for Content Provenance and Authenticity (C2PA) standards directly within camera sensors and audio capture hardware to cryptographically certify original capture integrity.

• **Modernize Statutory Liability Frameworks:** Legislative and banking authorities must modernize electronic fund transfer regulations so that victims targeted by sophisticated neural impersonation are not automatically disqualified from reimbursement under obsolete negligence standards.

• **Deploy Adaptive Liveness Biometrics:** Upgrade automated onboarding and remote access systems with active 3D flash reflection, randomized prompt interaction, and micro-

texture blood flow (photoplethysmography) analysis to defeat virtual camera injection streams.

AI-driven impersonation has evolved from simple text-based phishing into multi-modal social engineering that combines Large Language Models (LLMs), deepfake video synthesis, and real-time voice cloning.

Below is an analytical breakdown of landmark case studies, the technical dynamics of these attacks, and their grounding in contemporary AI ethics and cybersecurity frameworks.

1. Landmark Case Studies

The following real-world incidents illustrate how AI impersonation shifts cyber attacks from signature-based exploits to psychological and trust-based exploits:

Arup Corporate Fraud (Hong Kong, $25M Loss) | Multi-person Deepfake Video & Voice Cloning | Attackers recreated a video conference call featuring synthesized versions of the CFO and multiple colleagues, convincing a finance employee to execute 15 wire transfers. | Synthetic media can bypass multi-person visual confirmation protocols in real-time. |

| State-Sponsored & Dark LLM Extortion (Multinational Enterprise Targeting) | Jailbroken LLMs & Automated Reconnaissance | Threat actors used fine-tuned LLMs ("Dark LLMs") to parse public corporate data, draft hyper-personalized communications, and run parallel extortion campaigns. | AI reduces reconnaissance and social engineering timelines from weeks to minutes (a 100x+ speed multiplier). |

| IT Staffing & Supply Chain Infiltration (Remote Hiring Fraud) | Real-time Deepfake Filters & AI Identity Kits | Threat actors used generative AI to forge identity credentials, alter facial features during video interviews, and answer technical prompts for remote contractor roles. | Insider threat defense must account for AI-synthesized professional identities at the onboarding stage. |

2. Technical & Cybersecurity Challenges

[Reconnaissance] [AI Generation] [Multi-Channel Delivery] [Exploitation]

Public Data / Social > Voice Cloning & LLMs > Deepfake Call + Phishing > Unauthorized Transfer

* Failure of Human-centric Detection: Historically, users were trained to look for grammatical flaws, awkward phrasing, or unusual domains. Generative LLMs eliminate basic linguistic markers, producing context-aware, localized, and contextually fluent communications.

* Biometric Bypass: Voice cloning engines require under 5 seconds of sample audio to clone a voice with high tonal fidelity. This degrades traditional voice-biometric authentication used by financial institutions and enterprise IT helpdesks.

* Machine-Speed Scale: Attackers automate the entire attack lifecycle—from social media data scraping to message personalization—allowing targeted spear-phishing campaigns to scale exponentially.

3. Grounding in AI Ethics & Defensive Frameworks

Analyzing these attacks requires bridging regulatory frameworks with operational defense models:

Ethical & Regulatory Frameworks

* NIST AI Risk Management Framework (AI RMF): Emphasizes Trustworthiness—specifically Validity, Reliability, and Security. Impersonation directly attacks system validity by undermining user trust in digital communication layers.

* EU AI Act Requirements: Classifies biometric manipulation and synthetic media under high-risk and transparency obligations, mandating clear visual/audible watermarking (e.g., C2PA standards) for synthetic content.

* UNESCO Recommendations on the Ethics of AI: Focuses on Accountability and Privacy, highlighting the ethical duty of AI developers to prevent dual-use model weights (such as open-weight voice cloners) from being abused without safety rails.

Defensive Architectural Shift

* Zero Trust Architecture (ZTA): Under ZTA, identity is never assumed based on voice, video, or email context alone. Every critical request requires out-of-band verification via cryptographic multi-factor authentication (MFA) or pre-agreed secondary channels.

* Behavioral Anomaly Analytics: Shifting from static filtering to dynamic analysis—evaluating baseline communication patterns, request velocity, and unexpected authorization paths rather than media authenticity alone.

10. CONCLUSION

Alleviate the collapse of digital trust requires shifting from flawed post-hoc deepfake detection to systemic cryptographic origin, adaptive biometric authentication, and enforceable regulatory frameworks. As synthetic media becomes visually and acoustically indifferenciateable from genuine human interaction, uncorroborated audio and video can no longer serve as autonomous verification standards. Restoring integrity to the modern digital data demands a concerted fusion of zero-trust technical defenses, proactive institutional governance, and updated legal accountability frameworks that protect both enterprises and vulnerable citizens from synthetic deception.