INDIAN JOURNAL OF LEGAL REVIEW [IJLR – IF SCORE – 7.58]

VOLUME 6 AND ISSUE 12 OF 2026

APIS – 3920 - 0001 *(and)* ISSN - 2583-2344

Published by

Institute of Legal Education

https://iledu.in

MISUSE OF GENERATIVE AI IN CYBER CRIMES: DEEPFAKES, IDENTITY THEFT, AND ONLINE FRAUD

AUTHOR – TISHA . T, STUDENT AT CYBER SPACE LAW AND JUSTICE (SCHOOL OF EXCELLENCE IN LAW)

BEST CITATION – PRABHNOOR KAUR MAKKAR, MISUSE OF GENERATIVE AI IN CYBER CRIMES: DEEPFAKES, IDENTITY THEFT, AND ONLINE FRAUD, PROOF AND JUDICIAL RELIABILITY IN INDIA, *INDIAN JOURNAL OF LEGAL REVIEW (IJLR),* 6 (12) OF 2026, PG. 30-39, APIS – 3920 – 0001 & ISSN - 2583-2344.

Abstract

The rapid democratization of generative artificial intelligence (GenAI) has precipitated a structural transformation in the cyber threat landscape. Unlike traditional discriminative machine learning systems designed to classify or predict patterns, generative architectures—including Generative Adversarial Networks (GANs), diffusion models, neural audio codecs, and Large Language Models (LLMs)—synthesize high-fidelity text, audio, imagery, and video that are increasingly indistinguishable from authentic human output. This paper examines the weaponization of generative AI across three interconnected domains of cybercrime: synthetic media and deepfake exploitation, biometric and synthetic identity theft, and hyper-personalized online fraud. It analyzes how generative tooling collapses the marginal cost of sophisticated social engineering, bypasses legacy authentication safeguards, and facilitates "Crime-as-a-Service" ecosystems. From a legal and doctrinal perspective, the study evaluates the friction between emerging synthetic threats and existing statutory frameworks across the United States, the European Union, and India, while addressing constitutional constraints under free speech jurisprudence and evidentiary vulnerabilities in judicial proceedings, notably the "liar's dividend." Finally, the paper proposes an integrated governance framework combining cryptographic content provenance, heightened Know-Your-Customer (KYC) verification standards, and targeted statutory reforms to restore epistemic trust and accountability in digital ecosystems.

Keywords: Generative Artificial Intelligence, Deepfakes, Synthetic Identity Theft, Cyber Fraud, Social Engineering, Digital Forensics, Cyber Law

1. Introduction

The proliferation of generative artificial intelligence represents one of the most consequential technological inflection points of the twenty-first century. While foundation models and multimodal generative systems offer transformative gains in scientific discovery, software engineering, healthcare diagnostics, and creative industries, their dual-use archi-tecture inherently lowers barriers to malicious exploitation[103]. Historically, high-impactcyber operations required substantial technical expertise, bespoke infrastructure, and manual reconnaissance to craft convincing lures or forge authentic

[103] See generally Miles Brundage et al., The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation 16–20 (2018) (analyzing the dual-use nature of artificial intelligence systems and how capability gains expand the threat surface across digital, physical, and political security domains).

credentials. Generative AI disrupts this economic equilibrium by automating the production of deceptive artifacts at scale, enabling threat actors with modest technical proficiency to execute operations that previously demanded state-sponsored or organized syndicate resources[104].

At the core of this paradigm shift is the transition from analytical software to generative synthesis. Large Language Models (LLMs) generate contextually coherent, culturally nuanced, and grammatically flawless prose across dozens of languages; diffusion architectures and Generative Adversarial Networks (GANs) render photorealistic faces and realtime video overlays; and neural text-to-speech engines clone human vocal timbres from mere seconds of reference audio[105]. When operationalized by cybercriminals, these capabilities converge to undermine the foundational trust assumptions of digital communication: that a voice on a telephonecall belongs to the purported speaker, that a live video feed reflects physical presence, and that a well-crafted electronic message originates from a legitimate counterparty.

2.Technical Architecture of Malicious Generative AI

Understanding the legal and forensic challenges of AI-enabled cybercrime requires examining the mechanisms through which synthetic artifacts are produced. Modern generative pipelines rely on distinct neural architectures tailored to specific data modalities, each presenting unique exploitation pathways.

2.1 Core Generative Modalities

First, visual synthesis relies predominantly on Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and latent diffusion models. In a GAN framework, a generator network creates candidate images while a discriminator network evaluates their authenticity against real training distributions, iteratively refining output realism through zero-sum adversarial training[106]. Latent diffusion models, which have largely superseded GANs in high-resolution image and video generation, operate by iteratively denoising random Gaussian noise conditioned on text or image embeddings[107]. In cybercrime, autoencoders and real-time face-swapping pipelines extract latent facial expressions from a source actor and map them onto a target identity, enabling live impersonation during video conferencing sessions

Second, neural voice synthesis and zero-shot voice cloning utilize acoustic foundation models trained on vast corpora of human speech. By disentangling linguistic content from speaker identity embeddings, modern voice-cloning architectures replicate prosody, cadence, emotional inflection, and background acoustic characteristics from short audio samples harvested from social media posts, voicemail greetings, or public earnings calls[108].

[104] See Robert Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 Calif. L. Rev. 1753, 1758–62 (2019) (observing that machine-learning driven synthesis democratizes the capacity to fabricate hyper-realistic audiovisual media).
[105] See Ian J. Goodfellow et al., *Generative Adversarial Nets*, in 27 Advances in Neural Information Processing Systems 2672, 2672–75 (2014); Jonathan Ho et al., *Denoising Diffusion Probabilistic Models*, in 33 Advances in Neural Information Processing Systems 6840, 6840–42 (2020).
[106] Goodfellow et al., *supra* note 3, at 2673.
[107] See Robin Rombach et al., *High-Resolution Image Synthesis with Latent Diffusion Models*, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 10684, 10685–87 (2022).
[108] See Ye Jia et al., *Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis*, in 31 Advances in Neural Information Processing Systems 4480, 4481–83 (2018).

INDIAN JOURNAL OF LEGAL REVIEW [IJLR – IF SCORE – 7.58]
VOLUME 6 AND ISSUE 12 OF 2026
APIS – 3920 - 0001 *(and)* ISSN - 2583-2344
Published by
Institute of Legal Education
https://iledu.in

Third, transformer-based Large Language Models predict token sequences across vast context windows, enabling autonomous text generation, code synthesis, and interactive dialogue[109].Unlike static phishing templates, LLMs maintain stateful multi-turn conversations, dynamically adapting tone, psychological pressure, and technical vocabulary to mirror a victim's organizational context.

The Democratization of Offenses and "Crime-as-a-Service

Although frontier commercial models incorporate reinforcement learning from human feedback (RLHF) and safety guardrails designed to refuse harmful prompts, threat actors routinely circumvent these controls through adversarial "jailbreaking," prompt injection, or fine-tuning open-weight models with safety alignment layers removed[110].Furthermore, underground cybercrime marketplaces have commercialized malicious generative capbilities under a "Crime as-a-Service" (CaaS) model, offering subscription access to uncensored LLMs, turnkey deepfake webcam emulators, and automated voice-cloning APIs[111].

Model Architecture	**Primary Modality**	**Malicious Cybercrime Applications**
Autoencoders & GANs	Facial video & real time face swapping	Live KYC biometric bypass; executive video impersonation; non-consensual intimate imagery.
Latent Diffusion Models	High-resolution static & dynamic imagery	Forged identity documents; fabricated evidentiary photographs; fraudulent e-commerce listings.
Neural Audio / TTS Models	Zero-shot voice cloning & real-time voice conversion	Vishing (voice phishing); virtual kidnapping extortion; authorization of fraudulent wire transfers.
Large Language Models (LLMs)	Contextual text, multi-turn dialogue	Hyper-personalized spear-phishing; Busi

[109] See Ashish Vaswani et al., *Attention Is All You Need*, in 30 Advances in Neural Information Processing Systems 5998, 5999–6002 (2017).
[110] See Andy Zou et al., *Universal and Transferable Adversarial Attacks on Aligned Language Models*, arXiv:2307.15043, at 1–3 (2023) (demonstrating automated adversarial suffixes capable of bypassing alignment filters across aligned language models).
[111] See Europol, *ChatGPT: The Impact of Large Language Models on Law Enforcement* 4–7 (2023) (detailing how threat actors exploit LLMs for phishing, disinformation, and malware development).

	& source code	ness Email Compromise (BEC); polymorphic malware generation.

3 Typologies of Generative AI-Enabled Cyber Offenses

The operational deployment of generative AI spans a wide spectrum of criminal activity, reshaping established offense categories while creating novel attack surfaces.

3.1 Synthetic Media and Deepfake Exploitation

The term "deepfake"—a portmanteau of "deep learning" and "fake"—encompasses synthetic audio, image, or video content manipulated to depict individuals saying or doing things they never did[112]. While public discourse frequently focuses on electoral disinformation and geopolitical influence operations, the most pervasive harms involve targeted interpersonal and financial exploitation.

Non-consensual synthetic intimate imagery (NCII) represents the overwhelming majority of deepfake content circulating online.[113] Accessible nudification applications and fine-tuned diffusion checkpoints allow perpetrators to generate explicit depictions of private individuals using ordinary photographs scraped from social networking platforms. Beyond severe dignitary and psychological trauma, perpetrators weaponize synthetic intimate media for sexual extortion ("sextortion"), threatening to disseminate fabricated imagery to a victim's family, employer, or professional network unless ransom demands are mettives announcing product recalls, regulatory investigations, or bankruptcy filings can trigger algorithmic trading sell-offs within seconds, enabling perpetrators to profit through short-selling strategies before the fabrication is debunked.

3.2 Next-Generation Identity Theft and Biometric Spoofing

Identity theft has evolved from the unauthorized possession of static identifiers—such as Social Security numbers, Aadhaar numbers, or credit card details—to the dynamic synthesis and impersonation of a victim's biometric and behavioral persona. Generative AI accelerates two distinct forms of identity crime: synthetic identity fraud and biometric authentication bypass

In synthetic identity fraud, criminals combine real personally identifiable information (PII) obtained from data breaches with fabricated attributes to construct wholly fictitious personas.[114] Generative models automate the creation of consistent digital footprints for these synthetic identities, generating matching headshot portraits, forged governmentissued identification cards, utility bills, and employment histories to pass automated credit underwriting and open fraudulent financial accounts.

Even more disruptive is the use of real-time generative video and audio injection to defeat remote Know-Your-Customer (KYC) and biometric liveness checks. Financial institutions and cryptocurrency exchanges increasingly rely on facial recognition and active liveness challenges—requiring users to

[112] Chesney & Citron, *supra* note 2, at 1757
[113] See Danielle Keats Citron, *Sexual Privacy*, 128 Yale L.J. 1870, 1919–24 (2019) (examining how non consensual deepfake pornography inflicts severe psychological, reputational, and economic harm on victims)
[114] See Fed. Rsrv. Sys., *Synthetic Identity Fraud in the U.S. Payment System: A Review of Causes and Contributing Factors* 3–6 (2019).

blink, turn their heads, or read randomized digits aloud—to verify identity during remote onboarding.[115] Attackers circumvent these cotrols using virtual camera drivers that inject AI-driven face swapped video streams directly into the application layer, bypassing the physical optical sensor altogether. Coupled with real-time voice conversion, a single operator can simultaneously impersonate hundreds of distinct identity theft victims to drain bank accounts, hijack telecommunications SIM cards, and launder illicit proceeds.

3.3 Hyper-Personalized Online Fraud and Social Engineering

Online fraud schemes, particularly phishing and Business Email Compromise (BEC), have historically suffered from a trade-off between scale and personalization: mass-mailed lures were easily detected due to generic phrasing and linguistic errors, whereas be spoke spear-phishing required labor-intensive research. Large Language Models eliminate this constraint.[116] By ingesting open-source intelligence (OSINT) from corporate websites, LinkedIn profiles, and leaked email threads, LLMs craft context-aware lures referencing ongoing projects, internal jargon, and interpersonal relationships.

In Business Email Compromise and executive impersonation fraud, threat actors com bine LLM-generated correspondence with real-time audio and video deepfakes to authorize fraudulent wire transfers. In high-profile corporate incidents, finance personnel have been deceived into transferring tens of millions of dollars after participating in multi person video conference calls in which every other participant—including the purported Chief Financial Officer—was an AI-generated deepfake reconstructed from publicl available media.[117]

Additionally, generative AI assists fraudsters in building persuasive fraudulent storefronts, pig butchering romance-investment scams, and polymorphic malware. In longconinvestment frauds, conversational agents maintain continuous, emotionally engaing dialogues with targets across messaging platforms, overcoming language barriers and sustaining psychological manipulation over months before steering victims toward fraudulent trading portals.18 Simultaneously, code generating LLMs enable threat actors to obfuscate malware payloads and rewrite scripts dynamically to evade signature-based antivirus detection

4 Legal, Doctrinal, and Evidentiary Challenges

The rapid maturation of generative cybercrime exposes acute doctrinal tensions across substantive criminal law, intermediary liability, constitutional speech protections, and rules of evidence.

4.1 Attribution, Mens Rea, and Intermediary Liability

Establishing criminal liability requires proving both the prohibited conduct (*actus reus*) and the requisite culpable mental state (*mens rea*), anchored to an identifiable perpetrator. Generative AI

[115] See Nat'l Inst. of Standards & Tech. (NIST), *Digital Identity Guidelines: Enrollment and Identity Proofing*, NIST Special Pub. 800-63A, § 4.4 (2017) (establishing standards for remote identity proofing and biometric comparison)

[116] See Julian Hazell, *Spear Phishing with Large Language Models*, arXiv:2305.06972, at 2–5 (2023) (demonstrating that LLMs can generate convincing, personalized spear-phishing emails at minimal cost per target).

[117] See, e.g., Fin. Crimes Enf't Network (FinCEN), *FinCEN Alert on Fraud Schemes Involving Deepfake Media Targetng Financial Institutions*, FIN-2024-Alert004, at 2–4 (Nov. 13, 2024) (documenting typologies in which criminals utilize deepfake audio and video to bypass financial institution verification controls).

complicates attribution at multiple levels. First, synthetic artifacts lack the traditional physical provenance of forged documents or analog recordings, and threat actors frequently route generative API requests through layered anonymization networks and compromised cloud infrastructure. Second, as autonomous AI agents increasingly execute multi-step phishing, reconnaissance, and credential-stuffing workflows with minimal human supervision, prosecutors face doctrinal questions regarding foreseeability and proximate causation when an autonomous system deviates from or expands upon a user's initial prompt.

Equally contentious is the civil and regulatory liability of foundation model developers, hosting repositories, and social media platforms. In the United States, Section 230 of the Communications Decency Act shields providers of interactive computer services from liability as the publisher or speaker of information provided by another information content provider.[118]However, courts have long held that a platform loses Section 230 immunity to the extent it materially contributes to the development of unlawful content.[119] Whether the operator of a generative AI system acts as a neutral tool provider or a co-developer of unlawful synthetic media remains an unsettled frontier of statutory interpretation, particularly where models are specifically fine tuned or marketed for non-consensual faceswapping or voice cloning

4.2 Constitutional Boundaries and Free Speech Protections

Legislative efforts to criminalize or restrict the creation and distribution of deepfakes must navigate rigorous constitutional scrutiny, particularly under the First Amendment to the U.S. Constitution. In *United States v. Alvarez*, the Supreme Court held that false statements of fact are not categorically excluded from First Amendment protection merely because they are untrue; rather, content-based restrictions on false speech generally trigger strict scrutiny unless the speech falls within a historically recognized exception such as defamation, fraud, perjury, or true threats.

Consequently, broadly drafted statutes that prohibit all synthetic media without requiring proof of fraudulent intent, defamatory harm, or non-consensual intimate depiction risk invalidation on overbreadth grounds,[120] as they threaten to chill protected political satire, parody, artistic expression, and commentary. Conversely, statutes narrowly tailored to combat online fraud, identity theft, commercial impersonation, and non-consensual intimate imagery rest on firm constitutional footing, as fraudulent misrepresentation and severe invasions of sexual privacy do not enjoy First Amendment immunity[121]

4.3 Evidentiary Integrity and the "Liar's Dividend"

Beyond substantive offenses, generative AI threatens the integrity of judicial fact-finding. Under Rule 901(a) of the Federal Rules of Evidence, the proponent of an item of evidence must produce evidence sufficient to support a finding that the item is what the proponent claims it is Historically, audiovisual recordings carried substantial persuasive weight as objective witnesses. Today, however, opposing

[118] 47 U.S.C. § 230(c)(1).
[119] See *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1167–68 (9th Cir. 2008) (en banc).
[120] See *Hustler Magazine, Inc. v. Falwell*, 485 U.S. 46, 50–57 (1988) (affirming robust First Amendment protection for parody and satire involving public figures); see also Chesney & Citron, *supra* note 2, at 1791–95.
[121] See *Illinois ex rel. Madigan v. Telemarketing Assocs., Inc.*, 538 U.S. 600, 612 (2003) ("[T]he First Amendment does not shield fraud.").

parties can challenge authentic digital recordings as AI-generated fabrications, forcing courts to conduct costly forensic inquiries and rely on competing expert testimony under Rule 702 and the *Daubert* standard.

Scholars Chesney and Citron identify a perilous corollary to this phenomenon termed the "liar's dividend": as the public and fact-finders become increasingly aware that hyperrealistic audio and video can be fabricated, bad actors gain a powerful new defense mechanism to dismiss genuine recordings of their misconduct as deepfakes.[122]This epistemicerosion undermines both criminal prosecutions and civil accountability even in cases where no synthetic media was ever used.

5 Comparative Regulatory and Statutory Frameworks

Jurisdictions worldwide are adapting legacy cybercrime statutes while enacting specialized AI governance regimes to address synthetic fraud, identity theft, and deepfake abuse.

5.1 United States Federal and State Approaches

In the United States, federal prosecutors currently rely on technology-neutral criminal statutes to prosecute AI-enabled cybercrimes. Wire fraud under 18 U.S.C. § 1343 criminalizes any scheme or artifice to defraud executed via interstate electronic communications, fully encompassing LLM-driven phishing and deepfake BEC schemes.

Identity theft and aggravated identity theft under 18 U.S.C. §§ 1028 and 1028A prohibit the knowing transfer, possession, or use without lawful authority of a "means of identification" of another person—a definition broad enough to cover biometric identifiers and cloned vocal prints used in connection with predicate felonies.[123]Unauthorized access to computer systems facilitated by AI-assisted credential stuffing or malware falls under the Computer Fraud and Abuse Act (CFAA).[124]

Recognizing enforcement gaps in civil consumer protection, the Federal Trade Com

mission (FTC) promulgated the Rule on Impersonation of Government and Businesses, effective April 2024, which explicitly prohibits materially and falsely posing as a government entity or business and authorizes direct monetary penalties against scammers utilizing AI voice cloning and synthetic communications.[125] At the state level, jurisdictions including California, Texas, New York, and Virginia have enacted targeted statutes criminalizing nonconsensual deepfake intimate imagery, establishing rights of publicity against unauthorized digital replicas, and regulating deceptive synthetic media in electoral campaigns.[126]

5.2 The European Union AI Act and Digital Services Framework

The European Union has adopted a comprehensive, risk-based regulatory architecture anchored by the EU Artificial Intelligence Act (Regulation (EU) 2024/1689). Under Article 50 of the AI Act, providers of

[122] Chesney & Citron, *supra* note 2, at 1785–86.
[123] 18 U.S.C. §§ 1028(a)(7), 1028(d)(7), 1028A
[124]Rule on Impersonation of Government and Businesses, 16 C.F.R. pt. 461 (2024); see also Fed. Trade Comm'n, *FTC Proposes New Protections to Combat AI Impersonation of Individuals*, Press Release (Feb. 15, 2024) (proposing supplemental rules extending impersonation prohibitions to individual likenesses and voices).
[125]18 U.S.C. §§ 1028(a)(7), 1028(d)(7), 1028A.
[126] See, e.g., Cal. Penal Code § 647(j)(4) (criminalizing distribution of non-consensual intimate images); Va. Code Ann. § 18.2-386.2 (extending criminal penalties to falsified intimate video and images); Tenn. Code Ann. § 47-25-1101 *et seq.* (Ensuring Likeness, Voice, and Image Security (ELVIS) Act protecting voice rights against unauthorized AI cloning).

generative AI systems must ensure that synthetic audio, image, video, and text outputs are marked in a machine-readable format and detectable as artificially generated or manipulated. Furthermore, deployers of AI systems that generate or manipulate image, audio, or video content constituting a "deep fake" are legally obligated to disclose clearly that the content has been artificially generated or manipulated, subject to narrow exceptions for authorized law enforcement use or obvious artistic satire[127]

Complementing the AI Act, the Digital Services Act (DSA) requires Very Large Online Platforms (VLOPs) to assess and mitigate systemic risks stemming from the dissemination of illegal content and manipulative deepfakes, including rapid notice and-action mechanisms.38 Meanwhile, the General Data Protection Regulation (GDPR) classifies biometric data used for unique identification as a special category of personal data subject to strict processing prohibitions, providing individuals with erasure and objection rights against unauthorized scraping of their facial images and voice recordings.[128]

5.3 Comparative Perspectives: India and International Harmonization

In India, the rapid expansion of digital public infrastructure and mobile payments has made synthetic voice scams and deepfake impersonation an urgent enforcement priority. Prosecutors invoke several provisions of the Information Technology Act, 2000: Section 66C penalizes identity theft involving the fraudulent or dishonest use of electronic signatures, passwords, or unique identification features; Section 66D criminalizes cheating by personation using a computer resource; and Section 66E penalizes the violation of bodily privacy through the capture or transmission of private images. Additionally, the Bharatiya Nyaya Sanhita, 2023 (BNS), which replaced the Indian Penal Code, provides substantive offenses for cheating by personation, forgery of electronic records, defamation, and outraging the modesty of women, which courts apply to malicious deepfake deployments. The Ministry of Electronics and Information Technology (MeitY) has also issued advisories directing intermediaries under the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 to exercise due diligence in identifying and removing deepfakes and misinformation

At the international level, cross-border enforcement remains hampered by jurisdictional fragmentation. The Council of Europe's Budapest Convention on Cybercrime and its Second Additional Protocol facilitate expedited cross-border electronic evidence sharing,[129]while the United Nations Convention against Cybercrime seeks to establish harmonized global criminalization standards for computer-dependent and computer-enabled fraud.

[127] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), 2024 O.J. (L 2024/1689) [hereinafter EU AI Act].
[128] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation), arts. 9, 17, 2016 O.J. (L 119) 1, 38, 43–44.
[129] Convention on Cybercrime, Nov. 23, 2001, E.T.S. No. 185; Second Additional Protocol to the Convention on Cybercrime on Enhanced Co-operation and Disclosure of Electronic Evidence, May 12, 2022, C.E.T.S. No. 224.

Jurisdiction	Primary Instruments	Key Regulatory & Enforcement Focus
United States	18 U.S.C. §§ 1028, 1030, 1343; FTC Rule 16 C.F.R. pt. 461; State Deepfake Acts	Technology-neutral wire fraud and identity theft prosecution; civil penalties for AI impersonation; state-level NCII and voice rights protections.
European Union	EU AI Act (Reg. 2024/1689); Digital Services Act; GDPR	Mandatory machine-readable watermarking and deepfake disclosure (Art. 50); VLOP systemic risk mitigation; strict biometric data protection.
India	IT Act, 2000 (§§ 66C–66E); Bharatiya Nyaya Sanhita, 2023; IT Rules, 2021	Criminalization of cheating by personation via computer resources; intermediary due diligence and expedited takedown obligations.

6 Technical Countermeasures and Policy Recommendations

Addressing the systemic risks of generative AI in cybercrime requires a coordinated defensein-depth strategy integrating cryptographic[130] provenance, forensic detection, institutional verification reforms, and calibrated statutory updates.

[130] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, r. 3(1)(b) (India).

6.1 Cryptographic Provenance and Forensic Detection

Passive forensic detection—training machine learning classifiers to spot subtle visual or acoustic artifacts such as unnatural blinking,[131] spectral anomalies, or phase discontinuities—is locked in an asymmetric arms race. As generative architectures improve, forensic detectors suffer rapid accuracy degradation against unseen models and post-processing

Consequently, industry consortia and standards bodies are shifting emphasis toward

active cryptographic provenance. The Coalition for Content Provenance and Authentic

ity (C2PA) specification embeds tamper-evident, cryptographically signed metadat manifests directly into media files at the point of capture or generation. Rather than attempting to prove that a file is fake, cryptographic provenance establishes a verifiable chain of custody proving where, when, and by what hardware or generative model an asset was created or edited. Combined with imperceptible statistical watermarking embedded in model weights or token distributions, provenance standards offer a durable mechanism to distinguish authentic recordings from synthetic fabrications.

7 Conclusion

Generative artificial intelligence has fundamentally altered the mechanics of cybercrime, transforming deepfakes, identity theft, and online fraud from bespoke, labor-intensive operations into scalable, automated, and hyper-personalized campaigns. By synthesizing convincing human likenesses, voices, and contextual communications, malicious actors exploit the cognitive heuristics and institutional verification protocols upon which modern digital commerce and social trust depend.

Yet the weaponization of generative AI is neither technologically inevitable nor legally insurmountable. While existing fraud and identity theft statutes provide an essential baseline for prosecution, long-term resilience demands proactive alignment between technical architecture and legal governance. By pairing cryptographic content provenance and robust multi-factor authentication with targeted statutory prohibitions on malicious impersonation and non-consensual synthetic media, legal systems can preserve accountability, protect individual dignity, and safeguard the integrity of digital ecosystems in the generative era.

GRASP - EDUCATE - EVOLVE

[131] Convention on Cybercrime, Nov. 23, 2001, E.T.S. No. 185; Second Additional Protocol to the Convention on Cybercrime on Enhanced Co-operation and Disclosure of Electronic Evidence, May 12, 2022, C.E.T.S. No. 224