IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Mapping Criminal Harm in Generative AI: A Taxonomy-Based Legal Framework

Mansi Shukla, Dr Nidhi Arora

Ph.D Scholar, Assistant Professor

Banasthali Vidyapith, Department of Legal Studies

Abstract

The rapid diffusion of generative artificial intelligence (GenAI) has fundamentally altered the architecture of criminal harm by enabling scalable, low-cost, and highly realistic forms of digital misconduct. While existing legal frameworks nominally extend to such harms, they remain conceptually fragmented and doctrinally strained when applied to AI-mediated conduct. This article addresses this gap by developing a structured taxonomy of GenAI-enabled criminality, distinguishing between misinformation, fraud and identity theft, hate speech, and sexual offences, while recognizing deepfakes as a cross-cutting modality. Adopting a comparative doctrinal approach, the study analyses regulatory and criminal law responses in the United States, the European Union, and India. It argues that contemporary legal systems rely excessively on analogical extension of pre-existing offences. The article advances a taxonomy-driven framework for legal classification and proposes a recalibration of liability principles to address intent, agency, and harm in AI-mediated environments. It concludes that without doctrinal refinement and targeted regulatory intervention, criminal law risks both under-inclusivity and overreach in responding to synthetic harms.

1. Introduction

Generative artificial intelligence (GenAI) has rapidly transitioned from a specialised technological capability to a widely accessible infrastructure of content production. Systems capable of generating human-like text, synthetic audio, and hyper-realistic visual media are now embedded across social, economic, and political domains. This shift is not merely technological; it has significant implications for the structure and execution of criminal activity. By lowering barriers to entry and enabling automation at scale, generative systems have transformed both the means and the magnitude of harm. GenAI fundamentally alters the operational logic of traditional forms of digital crime such as fraud, defamation, identity theft, and dissemination of unlawful content. It enables the mass production of deceptive content, the personalization of manipulation strategies, and the creation of synthetic artefacts that are increasingly

indistinguishable from authentic human output. As a result, criminal conduct that once required coordination, expertise, or resources can now be executed by relatively low-skilled actors with minimal cost. ChatGPT by OpenAI, the most popular GenAI tool has numerous safety challenges ranging from Hallucinations; Harmful content; Disinformation and influence operations; Proliferation of conventional and unconventional weapons; Privacy; Cybersecurity; Potential for risky emergent behaviours, etc.[1] As per a report by U.S. government agency in 2024 Generative AI is being used to make highly potent tools to facilitate financial frauds. [2] This democratization of capability represents a qualitative shift in the criminogenic potential of digital technologies.

Despite this transformation, legal responses remain largely reactive and fragmented. Existing criminal law frameworks across jurisdictions, including the United States, the European Union, and India have primarily relied on the extension of pre-existing doctrines to AI-mediated harms, which in absence of institutional infrastructure fails to address deeper structural challenges introduced by generative systems, particularly with respect to attribution of intent, distribution of responsibility, and scale of harm. The absence of a coherent conceptual framework has resulted in inconsistent classification of offences, uneven enforcement, and uncertainty in liability attribution. A central difficulty lies in the mismatch between traditional legal assumptions and the operational characteristics of GenAI. Criminal law is premised on identifiable actors, intentional conduct, and reasonably foreseeable consequences. Generative systems, by contrast, function through probabilistic outputs, layered human and non-human inputs, and often unpredictable emergent behaviour. This creates doctrinal strain, particularly in distinguishing between intentional misuse and harmful outcomes arising from otherwise lawful interactions. Consequently, the question is no longer whether existing law can be applied to AI-enabled harms, but whether such application remains normatively coherent and practically effective.

This article addresses this gap by proposing a structured taxonomy of GenAI-enabled criminal harms, designed to align technological capabilities with legally recognizable categories of offence. It advances a four-fold classification comprising: (i) misinformation (ii) fraud and identity theft (iii) hate speech, and (iv) sexual offences, while treating deepfakes as a cross-cutting modality that complicates all four domains. The taxonomy is not merely descriptive; it is intended as an analytical tool to improve legal classification, clarify liability thresholds, and enable more consistent doctrinal application.

2. Methodology

This study adopts a comparative doctrinal research design to evaluate the adequacy of existing criminal law frameworks in addressing harms enabled by generative artificial intelligence (GenAI). The comparative inquiry is limited to the United States, the European Union, and India, chosen as

[1] OpenAI. (2023). GPT-4 System Card OpenAI (p. 44). https://cdn.openai.com/papers/gpt-4-system-card.pdf

[2] U.S. Department of the Treasury. (2024). Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Sector. In U.S. Department of the Treasury. U.S. Department of the Treasury. https://home.treasury.gov/news/press-releases/jy2212

representative regulatory archetypes. For the purpose of this analysis the primary legal materials used are statutes, judicial decisions, and regulatory instruments. The scope of the study is confined to four legally significant categories of harm: misinformation, fraud and identity theft, hate speech, and sexual offences. This taxonomy-based analytical framework, is constructed through inductive synthesis of recurring patterns of AI-enabled harm observed across legal and policy discourse. The taxonomy is designed to bridge the gap between technological capability and legal classification by organizing diffuse forms of misconduct into coherent, legally tractable categories. Its purpose is both analytical and normative: analytically, it provides a structured lens for examining doctrinal responses; normatively, it exposes limitations in existing frameworks and supports the argument for a more coherent and calibrated approach to criminal liability in the context of generative AI.

3. Conceptual Framework

To enable structured legal analysis, this section introduces a taxonomy that classifies GenAI-enabled harms into distinct, legally relevant categories. The taxonomy is designed to reduce conceptual ambiguity by grouping technologically diverse conduct according to the nature of harm and its engagement with criminal law principles. It serves as an organizing framework for the subsequent analysis and is as follows:

3.1. Taxonomy of GenAI-Enabled Harm

a. Misinformation

Misinformation is one of the major challenges imposed by GenAI.[3] Unlike conventional misinformation, which is typically human-authored and limited in dissemination capacity, GenAI enables automated production of context-specific and seemingly credible content, including synthetic text, audio, and deepfake media. Recent developments demonstrate its increasing deployment in politically sensitive contexts, including electoral processes in jurisdictions such as the United States and India, through the use of deepfake videos, automated bots, and micro-targeted messaging.

The regulatory responses to AI-enabled misinformation vary significantly across jurisdictions. In the United States, the legal approach is defined by a fragmented, sectoral regulatory framework strongly shaped by First Amendment protections, which prioritize the preservation of free expression over the proactive criminalization of synthetic content. While there is no unified federal legislation addressing AI-generated misinformation, several states have enacted targeted measures regulating the use of synthetic media in political campaigns. Most states, except Idaho, allow synthetic media subject to some safeguards like, disclosure requirements and temporal restrictions, labelling, etc, in election campaigns. Enforcement mechanisms range from civil remedies, such as injunctions and damages, to limited criminal sanctions in

[3] [3]Tang, Nan & Yang, Chenyu & Fan, Ju & Cao, Lei. (2023). VerifAI: Verified Generative AI. 10.48550/arXiv.2307.02796.

some states depending upon the use of media, under misdemeanours and felonies[4]. The 2024 deepfake robocall incident involving synthetic impersonation of a political figure (Joe Biden) illustrates the complexity of liability attribution, the proceedings are initiated under both Quasi-Judicial Action by (Federal Communications Commission) and Judicial Action by courts, the result of which (still pending) shall illustrate more in this context.

In India, misinformation is addressed through a combination of penal provisions and intermediary regulation. Statutory provisions under the Bharatiya Nyaya Sanhita, 2023 criminalize false statements in electoral contexts[5], public mischief[6], and reputational harm through forged electronic records[7], supplemented by criminal defamation provisions[8]. Parallelly, the Information Technology Act, 2000 and the Intermediary Guidelines and Digital Media Ethics Code Rules, 2021 impose due diligence obligations on intermediaries, including the removal of misleading content. However, the legal framework in India faces a transformative challenge regarding the application of "Safe Harbour" protections under Section 79 of the Information Technology Act, 2000. While Subsection 2(b)(ii) traditionally immunizes intermediaries that do not select the recipient of transmitted information, the advent of AI-driven micro-targeting disrupts this doctrinal assumption. In these automated environments, algorithmic systems actively determine the specific audience for content, thereby complicating the statutory requirement for neutrality. Consequently, judicial trends in India demonstrate an increasing reluctance to permit Section 79 to function as a blanket immunity for platforms in cases involving sophisticated technological mediation. [9] This shift necessitates a rigorous re-evaluation of intermediary liability, as the intersection of algorithmic amplification and generative synthesis may effectively void the "passive transmitter" defense. Following the 2023 amendments to Rule 3(1)(b)(v), intermediaries are mandated to exercise due diligence by ensuring users do not disseminate content that is intentionally deceptive or patently false. A critical component of this regime is the obligation for platforms to restrict content identified as misleading regarding Central Government business by an official Fact-Check Unit (FCU). Under Rule 3(1)(d), intermediaries must execute the removal of such flagged content within a 36-hour window upon receiving notice from judicial or sanctioned administrative authorities. The constitutional validity of these mandates was scrutinized in Kunal Kamra v. Union of India[10] before the Bombay High Court. The challenge raised several fundamental legal questions:

---

[4] Larkin, CJ. “Regulating Election Deepfakes: A Comparison of State Laws.” Tech Policy Press, 8 Jan. 2025, www.techpolicy.press/regulating-election-deepfakes-a-comparison-of-state-laws/. Accessed 8 Jan. 2026.

[5] Section 175 of the BNS, 2023

[6] Sec 353 of the BNS penalize wide scope of mischief from false information including in electronic medium.

[7] Section 336 (4) BNS

[8] Section 356 of the BNS

[9] There have been numerous cases like Christian Louboutin Sas vs Nakul Bajaj & Ors AIRONLINE 2018 DEL 1962, where the Delhi High Court made a distinction of actively involved intermediaries from others.

[10] MANU/MH/5903/2024

- Whether the FCU mechanism violates the principles of natural justice by positioning the government as an arbiter in its own cause.
- Whether "fake, false, or misleading" information constitutes a valid ground for speech restriction under the "reasonable restrictions" clause of Article 19(2).
- Whether the amendment is *ultra vires* by exceeding the delegated legislative powers granted under Sections 79 and 87 of the parent Act.
- Whether the lack of precise definitions for these terms renders the provision unconstitutionally vague, echoing the concerns addressed in the *Shreya Singhal* precedent.

The Bombay High Court delivered a split verdict; while the amendment was struck down by one justice for violating Articles 14 and 19, it was upheld by another, reflecting deep judicial disagreement over the limits of executive oversight in digital spaces. Currently, the broader constitutional validity of these rules remains a subject of significant litigation, with over 17 related petitions across various High Courts being consolidated [11] for a definitive resolution by the Supreme Court of India. [12]

The European Union adopts a predominantly regulatory approach, emphasizing risk mitigation rather than direct criminalization. Instruments such as the Digital Services Act and the AI Act (2024) impose transparency and labeling obligations for AI-generated content, thereby addressing misinformation through ex ante governance rather than ex post penal sanctions. Criminal liability continues to be governed by the domestic laws of Member States, where misinformation is prosecuted under established offences such as defamation, fraud, or incitement. For example; In France, the Law Against Manipulation of Information (2018) provides judicial mechanism to facilitate rapid interventions against the dissemination of false information during sensitive periods. Germany's Network Enforcement Act (NetzDG) imposes stringent compliance obligations on social media platforms with over two million users, mandating the deletion of illegal content within 24 hours and all other illegal material within seven days of a report. This statutory framework encompasses 22 distinct criminal categories, including defamation, incitement to violence, and public insult. Collectively, this reflects a broader tendency within European legal systems to subsume AI-enabled harms within existing doctrinal categories rather than creating technology-specific offences.

To conclude, from a doctrinal perspective, AI-enabled misinformation reveals a fundamental tension within criminal law: the difficulty of reconciling scale and automation with intent-based liability frameworks. While low-level harms remain within the domain of civil law, higher-order harms, particularly those affecting democratic processes, expose gaps in existing legal structures. Accordingly, misinformation in the context of GenAI may be more coherently understood across three axes: (i)

[11] SLP (C) No. 11163 of 2023.

[12] Ahuja, Ishwar, and Bhairavi S N. "An Analysis of Kunal Kamra vs Union of India." Bar and Bench - Indian Legal News, 7 Feb. 2024, www.barandbench.com/view-point/an-analysis-of-kunal-kamra-vs-union-of-india. Accessed 7 Jan. 2026.

individual harm, typically addressed through defamation and privacy law; (ii) institutional harm, particularly in the form of electoral manipulation; and (iii) systemic epistemic harm, arising from sustained exposure to distorted informational environments. Of these, the third category remains the least developed in legal doctrine, indicating a significant area for future regulatory and jurisprudential evolution.

b. Fraud/ Identity Theft

The integration of Generative AI into the landscape of financial crime has fundamentally augmented the efficacy of identity theft and social engineering. By utilizing high-fidelity facial and voice synthesis, actors can now bypass traditional biometric authentication systems and generate synthetic identification with unprecedented realism. This technological shift dramatically scales conventional digital misconduct, including business email compromise, phishing by enabling anonymous, automated impersonation that evades standard detection protocols. Consequently, GenAI does not merely introduce new harms but serves as a force multiplier for existing fraudulent modalities, elevating the complexity of digital deception. While most legal systems currently lack AI-specific penal codes for identity fraud, existing statutory frameworks provide a broad basis for prosecution. In the United States, the federal wire fraud statute (18 U.S.C. § 1343) [13] remains a primary enforcement tool due to its expansive scope, which encompasses any fraudulent scheme executed through electronic communications. Furthermore, Chapter 43 of the U.S. Code contains several specialized provisions targeting the false impersonation of specific entities, such as 18 U.S. Code section 911[14]; 18 U.S. Code section 912[15]; 18 U.S. Code section 913[16]; 18

[13] Wire Fraud, 18 U.S.C. § 1343 (2012). The section states, " Whoever, having devised or intending to devise any scheme or artifice to defraud, or for obtaining money or property by means of false or fraudulent pretenses, representations, or promises, transmits or causes to be transmitted by means of wire, radio, or television communication in interstate or foreign commerce, any writings, signs, signals, pictures, or sounds for the purpose of executing such scheme or artifice, shall be fined under this title or imprisoned not more than 20 years, or both. If the violation occurs in relation to, or involving any benefit authorized, transported, transmitted, transferred, disbursed, or paid in connection with, a presidentially declared major disaster or emergency (as those terms are defined in section 102 of the Robert T. Stafford Disaster Relief and Emergency Assistance Act (42 U.S.C. 5122)), or affects a financial institution, such person shall be fined not more than $1,000,000 or imprisoned not more than 30 years, or both."

[14] Impersonating Citizens of the United States: "Whoever falsely and willfully represents himself to be a citizen of the United States shall be fined under this title or imprisoned not more than three years, or both."

[15]Impersonating Officers or employees of the United States: "Whoever falsely assumes or pretends to be an officer or employee acting under the authority of the United States or any department, agency or officer thereof, and acts as such, or in such pretended character demands or obtains any money, paper, document, or thing of value, shall be fined under this title or imprisoned not more than three years, or both."

[16] Impersonators making arrest or search: "Whoever falsely represents himself to be an officer, agent, or employee of the United States, and in such assumed character arrests or detains any person or in any manner searches the person, buildings, or other property of any person, shall be fined under this title or imprisoned not more than three years, or both."

U.S. Code section 914[17]; 18 U.S. Code section 915;[18] 18 U.S. Code section 916[19]; 18 U.S. Code section 917[20]. At the sub-national level, certain jurisdictions have moved toward proactive legislative interventions to close the gap between technological advancement and legal response. For instance, New Hampshire recently enacted specific protections against the fraudulent use of deepfakes (RSA 638:26-a) [21], which establishes distinct criminal penalties and civil causes of action for synthetic impersonation. Such targeted measures reflect an emerging regulatory trend that seeks to calibrate liability principles specifically for the nuances of AI-mediated environments.

The European Union's strategy for mitigating the impact of AI-generated deepfake scams is structured around a tripartite framework of legislative, technological, and educational interventions. Legislative efforts prioritize transparency and data integrity through the following mechanisms:

- **Data Protection and Consent:** Under the General Data Protection Regulation (GDPR), entities are prohibited from collecting personal identifiers, including biometric voice data, without explicit consent. Furthermore, Article 22 mandates disclosure when AI systems are utilized for automated profiling or individual decision-making.
- **Algorithmic Transparency:** The EU AI Act (Article 50) requires providers to inform users of their interaction with synthetic content. It further stipulates that AI-generated or modified media must include machine-readable identifiers, such as digital watermarking, to ensure traceability.
- **Substantive Anti-Fraud Frameworks:** While various Member States have enacted specific anti-fraud legislation, criminal prosecution remains grounded in domestic penal codes that are increasingly being harmonized through regional directives.

[17] Impersonating Creditors of the United States: "Whoever falsely personates any true and lawful holder of any share or sum in the public stocks or debt of the United States, or any person entitled to any annuity, dividend, pension, wages, or other debt due from the United States, and, under color of such false personation, transfers or endeavors to transfer such public stock or any part thereof, or receives or endeavors to receive the money of such true and lawful holder thereof, or the money of any person really entitled to receive such annuity, dividend, pension, wages, or other debt, shall be fined under this title or imprisoned not more than five years, or both."

[18] Impersonating Foreign diplomats, consuls, or officers: "Whoever, with intent to defraud within the United States, falsely assumes or pretends to be a diplomatic, consular or other official of a foreign government duly accredited as such to the United States and acts as such, or in such pretended character, demands or obtains or attempts to obtain any money, paper, document, or other thing of value, shall be fined under this title or imprisoned not more than ten years, or both."

[19] Impersonating 4–H Club members or agents: "Whoever, falsely and with intent to defraud, holds himself out as or represents or pretends himself to be a member of, associated with, or an agent or representative for the 4–H clubs, an organization established by the Extension Service of the United States Department of Agriculture and the land grant colleges, shall be fined under this title or imprisoned not more than six months, or both."

[20]Impersonating Red Cross members or agents: "Whoever, within the United States, falsely or fraudulently holds himself out as or represents or pretends himself to be a member of or an agent for the American National Red Cross for the purpose of soliciting, collecting, or receiving money or material, shall be fined under this title or imprisoned not more than 5 years, or both."

[21] Fraudulent Use of Deepfakes, N.H. Rev. Stat. Ann. § 638:26-a (2024). https://legiscan.com/NH/text/HB1432/id/3002615

In Germany, fraud is prosecuted under Section 263 of the Strafgesetzbuch (StGB) [22], which necessitates proving deception, error, property disposition, and financial loss, coupled with the intent of unlawful enrichment. The French Code Pénal (Articles 313-1 to 313-3) [23] similarly criminalizes the fraudulent acquisition of property through deceptive practices or the abuse of trust. Additionally, the United Kingdom's Fraud Act 2006[24] provides a comprehensive basis for liability through three distinct categories: fraud by false representation, failure to disclose information, and abuse of position. These established doctrines are increasingly applied to AI-mediated environments to address the evolving nature of synthetic deception.

The Indian legal system addresses AI-facilitated fraud and identity theft through the combined application of the Information Technology Act, 2000 (IT Act) and the Bharatiya Nyaya Sanhita (BNS). Section 66C of the IT Act specifically criminalizes the fraudulent or dishonest use of another person's unique identification features, such as electronic signatures or passwords, imposing a penalty of up to three years' imprisonment and a fine of one lakh rupees. Furthermore, provisions regarding cheating by personation are established under both technological and general penal statutes. Section 66D of the IT Act imposes a term of up to three years and a fine for those who use a computer resource or communication device to cheat by personating another entity. Section 319 of the BNS broadens this scope by defining cheating by personation as pretending to be any other person, whether real or imaginary, and prescribes a more stringent punishment of up to five years' imprisonment. These statutory frameworks provide a robust, although technically general, basis for prosecuting digital impersonation crimes enabled by generative technologies.

c. Hate Speech

Hate speech can be defined as communications that incite discrimination, hostility, or violence against a group. Generative AI systems possess the capacity to synthesize and amplify hate speech. In contemporary digital ecosystems, such social harms are frequently escalated by automated bots that facilitate the rapid distribution of hate speech, example Microsoft's chatbot, Tay, deactivated in 2016 because it began disseminating racist content after interacting with biased user inputs.

The United States legal framework maintains a rigorous adherence to broad free speech protections, generally resisting pre-censorship and the categorization of hate speech as a specific criminal offense, though distinct forms of hate crimes are prosecuted under 18 U.S.C. § 249. The constitutional threshold for restricting such speech is governed by the Brandenburg Test (1969) [25] , which mandates that speech is

[22] Strafgesetzbuch [German Criminal Code], § 263 (1998). https://www.gesetze-im-internet.de/stgb/__263.html

[23] Code pénal [Penal Code], art. 313-1 to 313-3 (2002). https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000006418192

[24] Fraud Act 2006 (United Kingdom), c. 35. https://www.legislation.gov.uk/ukpga/2006/35/contents

[25] Brandenburg v. Ohio, 395 U.S. 444 (1969).

only punishable if it is directed toward inciting, and is likely to produce imminent lawless action. Consequently, AI-generated slurs or rhetoric, regardless of their nature, typically remain protected under the First Amendment unless they transition into a direct incitement of illegal acts. This creates a significant doctrinal challenge where synthetic hate speech may fall outside the reach of criminal law despite its potential for social destabilization.

The European Union's strategy for addressing AI-mediated hate speech integrates substantive criminal prohibitions with robust regulatory oversight. This framework is anchored by the Council Framework Decision 2008/913/JHA[26], which mandates that Member States criminalize the intentional public incitement of violence or hatred directed at groups based on race, religion, or ethnicity. Furthermore, Article 1 of this Decision requires the penalization of deliberate conduct to provoke hostility toward racial, religious, or ethnic communities, or a conduct that publicly condones, denies, or grossly trivializes international crimes such as genocide or war crimes, while Article 4 of this decision obligates member states to criminilise racist and xenophobic motivations. In contrast to the permissive standards in the United States, several European jurisdictions maintain more restrictive speech doctrines. For instance, Germany's Strafgesetzbuch (§ 130)[27] prescribes imprisonment ranging from three months to five years for acts of incitement, while the United Kingdom's Public Order Act 1986 (Part III) [28] provides a comprehensive basis for prosecuting the use of threatening or insulting words intended to stir up racial hatred. Beyond traditional penal codes, the EU employs contemporary regulatory instruments to manage algorithmic harms. The Digital Services Act (2022)[29] seeks to establish a safe online environment by institutionalizing content flagging, mandatory complaint mechanisms, and transparency reports for digital platforms. These measures, complemented by the European Commission's various codes of conduct, reflect a normative stance that views synthetic hate speech not merely as a technological byproduct, but as a systemic threat to fundamental human rights and social cohesion.

In the Indian constitutional framework, the right to free expression under Article 19(1)(a) is subject to specific reasonable restrictions, including the preservation of sovereignty, state security, public order, and morality. Consequently, AI-generated content is only subject to criminal sanctions if it intersects with these established exceptions. Under the Bharatiya Nyaya Sanhita (BNS), hate speech is not criminalized per se but is penalized when it threatens public tranquility; specifically, Section 196 targets the promotion of enmity between diverse social groups, while Sections 299 and 302 address the intentional wounding of religious sentiments. Complementing these substantive penal provisions, Section 69A of the Information

[26] Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law, Official Journal of the European Union, L 328/55 (2008). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32008F0913

[27] Strafgesetzbuch [German Criminal Code], § 130 (1998). https://www.gesetze-im-internet.de/stgb/__130.html

[28] Public Order Act 1986 (c. 64), s. 18. https://www.legislation.gov.uk/ukpga/1986/64/section/18

[29] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), Official Journal of the European Union, L 277/1 (2022). https://eur-lex.europa.eu/eli/reg/2022/2065/oj

Technology Act, 2000, which survived the judicial scrutiny in Shreya Singhal v Union of India[30], empowers the executive to direct the removal of digital content that jeopardizes national integrity or incites cognizable offenses. Together, these statutes form a reactive governance model that relies on traditional public order triggers to regulate the output of generative systems.

d. Sexual Offences

The emergence of non-consensual synthetic pornography, often termed "deep nudes," represents the most pervasive and gendered risk within the generative AI ecosystem.[31] Empirical data from 2023 indicates that pornographic content constitutes approximately 98% of all online deepfake videos, with women targeted in 99% of identified cases.[32] The accessibility of this technology is a critical driver of harm; research suggests that one-third of available deepfake tools explicitly facilitate the creation of sexually explicit material, often requiring less than 25 minutes of processing time and zero financial cost to transform a single static image into a 60-second synthetic video.[33]

The escalation of AI-generated Child Sexual Abuse Material (CSAM) further underscores the severity of this crisis. Reports from the Internet Watch Foundation (an update in 2024[34] on its 2023[35]) noted the discovery of 3,500 new synthetic CSAM images on dark web forums in the ten-month period leading up to July 2024 alone. Global controversies involving high-traffic platforms and specialized applications, such as those enabling users to virtually undress individuals for a nominal fee or no fee have demonstrated the rapid normalization of these harms. Incidents involving the creation and peer-to-peer distribution of synthetic nudes among students in Spain[36] and the United States[37], highlight a disturbing shift in the nature of sexual violence. Unlike traditional non-consensual pornography, which often required physical access or coercion, generative AI allows for the automated manufacturing of abuse from publicly available data. The easy accessibility of this technology necessitates an urgent legislative response to address the systemic victimization of women and minors in the digital age.

---

[30] AIR 2015 SC 1523

[31] Franks, M., & Waldman, A. (2019). Maryland Law Review Sex, Lies, and Videotape: Deep Fakes and Free Speech Delusions. Maryland Law Review, 78(4), 892–898. https://digitalcommons.law.umaryland.edu/cgi/viewcontent.cgi?article=3835&context=mlr

[32] Security Hero. (2023). 2023 state of deepfakes: Realities, threats, and impact. https://www.securityhero.io/state-of-deepfakes/

[33] Ibid

[34] IWF (Internet Watch Foundation). (2024). What has changed in the AI CSAM landscape? In www.iwf.org.uk. IWF. https://admin.iwf.org.uk/media/nadlcb1z/iwf-ai-csam-report_update-public-jul24v13.pdf

[35] IWF (Internet Watch Foundation). (2023). How AI is being abused to create child sexual abuse imagery. In www.iwf.org.uk. IWF. https://admin.iwf.org.uk/media/nadlcb1z/iwf-ai-csam-report_update-public-jul24v13.pdf

[36] Safi, M., Atack, A., & Kelly, J. (2024, February 29). Revealed: The names linked to ClothOff, the deepfake pornography app. The Guardian. https://www.theguardian.com/technology/2024/feb/29/clothoff-deepfake-ai-pornography-app-names-linked-revealed

[37] Yale Law School. (2024, October 31). *Clinics file suit against website that generates nonconsensual nude images*. https://law.yale.edu/yls-today/news/clinics-file-suit-against-website-generates-nonconsensual-nude-images

In the United States, the federal response to non-consensual synthetic pornography is anchored in 18 U.S.C. §§ 2252-2252A, which criminalize the production and distribution of child sexual abuse material (CSAM). While federal legislative proposals specifically targeting adult non-consensual deepfakes have gained significant momentum throughout 2024 focusing on criminalizing the synthesis of intimate content and imposing mandatory takedown protocols for digital platforms, the primary engine of legal evolution remained at the state level. By late 2024, state legislatures increasingly moved to close the analog gap in their penal codes by expanding the definition of disorderly conduct and harassment to encompass AI-generated media. California emerged as a regulatory leader through two pivotal 2024 legislative actions: Senate Bill 926 to amend Section 647 of the Penal Code, to Division 8 of the Business and Professions Code, relating to social media platforms. SB Senate Bill 926 amended Section 647 of the California Penal Code to categorize the creation and distribution of non-consensual intimate content, whether authentic or AI-generated, as a misdemeanor, punishable by imprisonment up to six months and fine up to $1,000 for first offense and imprisonment up to one year and fine up to $2,000 for subsequent offense. Senate Bill 981 introduced a mandate requiring social media providers to institutionalize reporting mechanisms for sexually explicit digital identity theft. This shift in the Business and Professions Code effectively transitions the burden of immediate content mitigation from the victim to the intermediary, emphasizing the necessity of rapid response in preventing the viral spread of synthetic harms. By late 2024, these state frameworks established a bifurcated liability model that differentiates between misdemeanors for simple distribution and felonies for cases involving significant reputational or emotional damage. This decentralized evolution reflects a broader shift in American jurisprudence: the move away from viewing deepfakes as mere privacy violations toward treating them as specialized forms of digital identity theft and sexual aggression.

The European Union has established a robust normative trajectory for addressing synthetic sexual exploitation through Directive (EU) 2024/1385[38], enacted in May 2024. This instrument mandates that Member States criminalize the intentional synthesis, modification, or public dissemination of non-consensual pornographic material facilitated by Information and Communication Technologies (ICT). Specifically, Article 5(1)(b) targets conduct likely to cause serious harm to the depicted individual, while Article 5(2) extends criminal liability to the use of such synthetic media as a tool for coercion or extortion. While the directive provides a uniform mandate, the subjective interpretation of the serious harm threshold remains a point of doctrinal debate, potentially leading to divergent enforcement standards across different jurisdictions during the transposition period. Prior to the full integration of these criminal provisions by mid-2027, the EU relies on a precautionary regulatory framework anchored by the Digital Services Act (DSA) and the AI Act. Article 35 of the DSA requires digital platforms to implement clear labeling for deepfakes, while Article 50 of the AI Act imposes a technical mandate on AI deployers to ensure that all synthetic content is marked in a machine-readable format. This layered approach ensures that while

[38] European Parliament and Council of the European Union. (2024, May 14). Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on combating violence against women and domestic violence. Official Journal of the European Union, L, 2024/1385.

substantive criminal law evolves to address the specificities of AI-mediated violence, ex-ante transparency measures serve as an immediate technical safeguard against the proliferation of non-consensual intimate imagery.

In the Indian legal framework for the prosecution of non-consensual synthetic pornography and AI-generated child sexual abuse material is primarily grounded in the Information Technology (IT) Act, 2000, and the Protection of Children from Sexual Offences (POCSO) Act, 2012. Section 67A of the IT Act provides a broad statutory basis for criminalizing the publication or transmission of sexually explicit material in electronic form. Its technologically neutral wording encompasses both authentic and AI-synthesized content, regardless of consent, prescribing up to five years of imprisonment and with fine which may extend to ten lakh rupees and for subsequent conviction imprisonment of either description for a term which may extend to seven years and also with fine which may extend to ten lakh rupees. In addition to this, Section 67B specifically targets the electronic distribution of CSAM, with punishment on first conviction with imprisonment of either description for a term which may extend to five years and with fine which may extend to ten lakh rupees and in the event of second or subsequent conviction with imprisonment of either description for a term which may extend to seven years and also with fine which may extend to ten lakh rupees. Section 67 imposes punishment of imprisonment of either description for a term which may extend to three years and with fine which may extend to five lakh rupees for conviction and in the event of second or subsequent conviction with imprisonment of either description for a term which may extend to five years and also with fine which may extend to ten lakh rupees for transmission of obscene material. Furthermore, Section 66E addresses the foundational harm of privacy invasion by penalizing the unauthorized publication of images depicting an individual's private areas. The protection of minors is further reinforced by the POCSO Act[39], which adopts a comprehensive definition of the use of a child for pornographic purposes [40]. Under Section 13, the act of utilizing a minor in any media format, including internet-based or synthetic representations, for sexual gratification is criminalized. Section 14 establishes a graduated sentencing matrix as follows:

- Using children for Pornographic purpose:
  - First Offense: Imprisonment for up to 5 years and a fine.
  - Subsequent Offenses: Imprisonment for up to 7 years and a fine.
- Direct Participation in Pornographic Acts: If the offender directly participates in the pornographic acts while using a child, the punishments are as follows for:
  - Penetrative Sexual Assault: Imprisonment for 10 years to life plus a fine.

[39] Protection of Children from Sexual Offences Act, 2012, No. 32, Acts of Parliament, 2012 (India). https://www.indiacode.nic.in/bitstream/123456789/2079/1/AA2012-32.pdf

[40] The section states, "—Whoever, uses a child in any form of media (including programme or advertisement telecast by television channels or internet or any other electronic form or printed form, whether or not such programme or advertisement is intended for personal use or for distribution), for the purposes of sexual gratification, which includes— (a) representation of the sexual organs of a child; (b) usage of a child engaged in real or simulated sexual acts (with or without penetration); (c) the indecent or obscene representation of a child, shall be guilty of the offence of using a child for pornographic purposes."

- Aggravated Penetrative Sexual Assault: Mandatory rigorous life imprisonment for and a fine.
- Sexual Assault: Imprisonment not less than 6 years but extend up to 8 years and a fine.
- Section 9 (Aggravated Sexual Assault): Imprisonment not less than 8 years but extend up to 10 years and a fine.

4. Comparative Synthesis and Critical Analysis

The preceding analysis reveals that, across jurisdictions, legal systems have largely responded to GenAI-enabled harms through doctrinal extension rather than structural adaptation. While existing offences such as fraud, defamation, and electoral misconduct remain formally applicable, their underlying assumptions are increasingly strained when confronted with the scale, automation, and diffusion of agency characteristic of generative systems. This results in a pattern of partial legal adequacy, where low-intensity harms are effectively addressed, but higher-order, systemic harms remain insufficiently captured.

A central point of convergence across the United States, the European Union, and India is the reliance on technology-neutral legal frameworks. In all three systems, AI-enabled conduct is subsumed within pre-existing categories of criminal and civil liability. However, the effectiveness of this approach varies depending on the nature of harm and the effectiveness of the technological and forensic infrastructure for law enforcement. In cases such as fraud and identity theft, where deception, intent, and harm can be more clearly established, traditional doctrines continue to function with relative coherence. By contrast, in domains such as misinformation and algorithmic amplification, where harm is diffuse and intent is difficult to attribute (in cases of filter bubble, etc), legal responses become fragmented and often inconsistent. Criminal law is premised on identifiable actors and intentional conduct, yet GenAI systems operate through layered interactions involving developers, deployers, and end-users. This diffusion of agency complicates the assignment of culpability, particularly in cases of dangerous or unintended harm. The comparative analysis demonstrates that none of the examined jurisdictions has developed a fully coherent framework to address this challenge. Instead, liability is either narrowly imposed on direct actors, risking under-inclusivity or broadly extended to intermediaries, raising concerns of overreach and chilling effects.

Another recurring limitation lies in the problem of scale. Generative AI enables the rapid and widespread dissemination of harmful content, transforming isolated acts into mass phenomena. Legal frameworks, however, remain oriented toward individualized harm and case-by-case adjudication, making them ill-equipped to respond to systemic risks such as large-scale misinformation campaigns or algorithmically reinforced echo chambers. While the European Union attempts to address this through ex ante regulatory obligations, particularly under platform governance regimes, such measures operate alongside, rather than within, traditional criminal law structures.

The comparative divergence is most visible in the normative prioritization of competing values. The United States adopts a speech-protective approach, limiting the scope of criminalization even in the face of technologically amplified harms. The European Union emphasizes risk mitigation and platform accountability, privileging preventive regulation over punitive enforcement. India, in contrast, reflects a

hybrid model that combines broad penal provisions with intermediary regulation, but faces ongoing constitutional tensions regarding overbreadth, vagueness, and state control over information. These differences highlight that legal responses to GenAI-enabled harm are not merely technical, but deeply shaped by constitutional and institutional contexts.

Taken together, these findings indicate that the current legal landscape is characterized by reactive adaptation, doctrinal fragmentation, and normative inconsistency. While existing laws provide a necessary foundation, they are insufficient to fully address the distinctive challenges posed by generative AI in the absence of advance technically backed infrastructure for forensics and law enforcement.

5. Recommendations

Based on the taxonomy-driven legal framework and the comparative analysis of the United States, European Union, and India, the following recommendations are proposed to address the doctrinal strain and regulatory gaps identified in the context of generative AI-enabled crimes.

- For successful implementation of the existing laws on the new genus of offences the infrastructural development is needed and R&D investments is required in technologies that can combat such harms.
- The traditional passive transmitter defense for intermediaries must be re-evaluated.
- Safe Harbour protections, such as those under Section 79 of the IT Act, should be voided when algorithmic systems actively engage in micro-targeting or the amplification of synthetic harms, as these actions determine the content's audience and disrupt statutory neutrality.
- Mandatory Takedown Protocols should be ensured, with expedited removal windows, ranging from 24 to 48 hours, should be institutionalized for non-consensual intimate content and misleading information regarding core democratic processes.
- Social media platforms should be legally required to establish specialized reporting channels for "sexually explicit digital identity theft," shifting the burden of mitigation from the victim to the intermediary.
- Regulatory frameworks should mandate that all AI deployers incorporate machine-readable identifiers and digital watermarking in synthetic outputs to ensure traceability and transparency.
- Liability principles should shift from focusing solely on content to conduct, I.e; there should be appropriate provisions to penalise the deliberate deployment of automated systems known to produce harmful or discriminatory outputs.
- Legal frameworks should specifically criminalize the *creation* of non-consensual synthetic pornography, acknowledging that GenAI allows for the automated manufacturing of abuse without the physical coercion required in traditional crimes.
- Given the rapid escalation of AI-generated Child Sexual Abuse Material (CSAM), jurisdictions must reinforce statutes like the POCSO Act and 18 U.S.C. §§ 2252-2252A to ensure they explicitly cover synthetic representations of minors used for sexual gratification.
- States should implement temporal restrictions and disclosure requirements for synthetic media used in political campaigns to safeguard democratic institutional integrity.

6. Conclusion

The taxonomy developed in this study demonstrates that different categories of GenAI-enabled harm engage legal principles in materially distinct ways, thereby underscoring the limits of a uniform or technology-neutral approach. As demonstrated through the proposed taxonomy, GenAI functions as a

force multiplier for traditional offenses, including misinformation, fraud, hate speech, and sexual violence, while introducing unique challenges to the operational logic of criminal law. The central claim advanced in this paper is that the problem is not the absence of legal tools, but their misalignment with the operational logic of generative systems. Criminal law continues to be anchored in assumptions of identifiable actors and discrete acts, whereas GenAI enables decentralized and probabilistic forms of conduct. Without conceptual restructuring, this misalignment risks producing both under-enforcement of serious harms and over-expansion of liability in ways that threaten fundamental rights. The proposed taxonomy-driven framework offers a structured path for legal classification and a necessary recalibration of liability principles to address intent, agency, and scale in AI-mediated environments. To safeguard the digital ecosystem, the legal focus must transition from merely regulating content to addressing the systemic risks inherent in the deliberate deployment of automated generative systems.