DEEP FAKES AND INDIAN CRIMINAL LAW: CHALLENGES AND REGULATING GAPS IN THE AGE OF AI

Alisha Chauhan, Phoolwati College, CCSU

ABSTRACT

The proliferation of Generative Artificial Intelligence (GenAI) has weaponized digital media through "deepfakes"—highly convincing, algorithmically manipulated audio, visual, or audio-visual content. In the Indian context, the malicious application of this technology spans across critical dimensions: financial extortion, electoral subversion, targeted political disinformation, and the involuntary sexualization of individuals through Non-Consensual Intimate Imagery (NCII). As these synthetic fabrications blur the boundary between fiction and reality, they pose an unprecedented threat to individual privacy, bodily autonomy, institutional trust, and national security.

In light of these challenges, this study evaluates the efficacy of India's current penal and digital architectures. Specifically, it investigates whether the newly enforced Bharatiya Nyaya Sanhita, 2023 (BNS) and the Information Technology Act, 2000 (IT Act) adequately address AI-generated crimes, and to what extent the IT Intermediary Guidelines successfully regulate deepfakes without causing a chilling effect on free speech. Adopting a doctrinal research methodology, this paper analyzes statutory provisions, global risk-based governance models like the EU AI Act, and emerging judicial precedents, including personality rights injunctions by various High Courts.

The findings reveal that a fragmented "patchwork approach" persists within Indian criminal law; the BNS lacks a distinct criminalization of AI offenses, forcing reliance on traditional definitions of fraud or defamation. Furthermore, recent regulatory updates mandating strict takedown windows for Synthetically Generated Information (SGI) shift an immense burden onto intermediaries, risking defensive over-censorship. The paper concludes that algorithmic accountability cannot serve as a permanent proxy for substantive criminal reform. It recommends amending the BNS to codify a standalone, technology-sensitive offense penalizing malicious synthetic media, establishing specialized digital forensics infrastructure, and formulating clearer statutory defenses to shield bona fide parodies and research.

Keywords: Bharatiya Nyaya Sanhita (BNS) 2023, Deepfakes, Digital Privacy, Generative AI (GenAI), Information Technology Act 2000, Intermediary Liability, Non-Consensual Intimate Imagery (NCII), Synthetically Generated Information (SGI)

1.1 Conceptual Background and Technological Evolution

The dawn of the Fourth Industrial Revolution has transitioned human civilization into an algorithmic paradigm, wherein Artificial Intelligence (AI) serves as both an engine of innovation and an instrument of sophisticated malice. Among the most potent and disruptive manifestations of Generative Artificial Intelligence (GenAI) is the creation of "deepfakes"—hyper-realistic, synthetically manipulated digital artifacts that superimpose or alter audio, visual, or audio-visual data using advanced Deep Learning models, primarily Generative Adversarial Networks (GANs). While synthetic media holds immense potential for creative industries, its unregulated weaponization poses structural challenges to sovereign legal systems. In India, the malicious application of deepfakes has rapidly escalated from isolated technological anomalies to systemic threats encompassing digital financial fraud, targeted political disinformation during democratic exercises, and severe violations of gender dignity through the automated generation of Non-Consensual Intimate Imagery (NCII).

1.2 Constitutional Conundrum: Privacy and Bodily Autonomy

The constitutional subtext of this crisis intersects directly with the fundamental right to privacy and digital integrity guaranteed under Article 21 of the Constitution of India. As the Supreme Court of India observed in the landmark judgment of ***K.S. Puttaswamy v. Union of India (2017) 10 SCC 1***, informational privacy, cognitive control, and dominion over one's digital persona are intrinsic to human dignity and bodily autonomy. Deepfakes fundamentally violate this constitutional ethos by stripping individuals of consent over their own likeness and voice, allowing external algorithms to hijack their physical and verbal identities.

1.3 The Contemporary Legislative Matrix in India

To counter this surging digital threat, India's criminal justice system underwent a massive statutory overhaul with the implementation of the **Bharatiya Nyaya Sanhita, 2023 (BNS)**, which repealed the colonial-era Indian Penal Code (IPC). The BNS, read alongside the **Information Technology Act, 2000 (IT Act)**, forms the primary penal framework

deployed by law enforcement to address these technological transgressions. For instance, prosecutors frequently look toward Section 356 of the BNS for criminal defamation, or Sections 66C (Identity Theft) and 66D (Cheating by Impersonation) of the IT Act to penalize deepfake-driven financial scams.

However, these statutes operate on traditional thresholds of criminality—such as proving physical proximity, structural forgery, or specific financial "dishonesty"—which often fail to capture the nuances of algorithmically generated digital harms that cause profound psychological, social, and reputational havoc without a direct monetary nexus.

1.4 The Regulatory Frontier and Platform Governance

Simultaneously, the regulatory frontier has shifted toward platform governance via the **Ministry of Electronics and Information Technology (MeitY)**. Recognizing the immediate threat of synthetic manipulation to public order and democratic stability, MeitY has enforced strict directives under the **Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules**. These guidelines mandate that social media intermediaries actively identify and purge "Synthetically Generated Information" (SGI) within strict, time-bound windows.

While this timeline-centric enforcement attempts to throttle the viral spread of deepfakes, it sparks critical constitutional questions regarding free speech. As articulated by the Supreme Court in ***Shreya Singhal v. Union of India (2015) 5 SCC 1***, intermediary liability frameworks must not lead to private platforms acting as self-appointed, defensive censors of online expression, which could trigger a "chilling effect" on legitimate digital speech, political parody, and artistic satire under Article 19(1)(a).

THE STATUTORY LANDSCAPE — LEGAL FRAMEWORKS AND CORE SECTIONS

The prosecution of deepfake-related offenses in India requires a complex synchronization between substantive criminal law, specialized cyber legislation, and electronic evidentiary rules. Because the Indian legislature has not yet enacted a standalone "Anti-Deepfake Act," law enforcement agencies rely on specific sections scattered across three core statutes. This chapter maps out the explicit legal frameworks and statutory provisions applicable to synthetic

media manipulation.

1. The Bharatiya Nyaya Sanhita (BNS), 2023

The BNS serves as the primary penal tool for punishing the criminal intent (*mens rea*) and social consequences of deepfakes, primarily focusing on reputational, gender-based, and state-centric offenses.

Section 356 (Criminal Defamation): This section replaces Section 499 of the IPC. It criminalizes the making or publishing of any "visible representation" intended to harm a person's reputation. Deepfakes, by creating hyper-realistic false depictions of individuals saying or doing things they never did, constitute a direct violation under this section.

Section 336 & 340 (Forgery and Making a False Electronic Record): These provisions penalize the creation of a "false electronic record" with the intent to cause damage or fraud. Altering video pixels or audio frequencies via Generative Adversarial Networks (GANs) to deceive the public is prosecuted under these sections.

Section 77 (Voyeurism) & Section 79 (Insulting the Modesty of a Woman): These gender-specific provisions are invoked against Non-Consensual Intimate Imagery (NCII). Section 79 specifically targets any word, gesture, or act (including digital images) intended to insult a woman’s modesty, making it the primary tool against deepfake pornography.

Section 152 (Act endangering sovereignty, unity, and integrity of India) & Section 353 (Public Mischief): Deployed when political deepfakes or cloned audios are virally circulated to incite riots, disturb public tranquility, or disrupt democratic elections.

2. The Information Technology (IT) Act, 2000

The IT Act provides the specialized technological framework required to address crimes committed strictly within cyberspace and governs electronic data manipulation.

Section 66C (Identity Theft): This section penalizes the fraudulent use of any "unique identification feature" of a person. In high-tech deepfakes, an individual’s biometric facial structure or unique vocal frequency clone is classified as a digital identification feature, bringing voice-cloning scams under this domain.

Section 66D (Cheating by Impersonation using Computer Resource): Explicitly punishes anyone who cheats by impersonation using a communication device. This section is heavily utilized in real-time FinTech scams where fraudsters use real-time video/audio deepfakes to impersonate corporate heads or family members.

Section 67 & 67A (Transmission of Obscene and Explicit Material): These sections penalize the electronic publication and transmission of sexually obscene or explicit material. They carry strict, non-bailable punishments for distributing deepfake pornography over the internet.

3. The Bharatiya Sakshya Adhiniyam (BSA), 2023

The BSA (which replaced the Indian Evidence Act, 1872) regulates how deepfakes are presented and authenticated as electronic evidence in a court of law.

Section 61 & 63 (Admissibility of Electronic Records): These provisions dictate the strict legal procedure for proving the authenticity of digital data. For a deepfake video to be legally analyzed or used by the prosecution, it must be accompanied by a mandatory statutory certificate under Section 63, verifying the integrity of the computer device and the hash value of the media file.

JUDICIAL PRECEDENTS — EVALUATING THE INTERPRETATIVE APPROACH OF INDIAN COURTS

In the absence of explicit, standalone legislation targeting artificial intelligence and synthetic media, the Indian judiciary has emerged as the primary institutional bulwark against deepfake offenses. By relying on creative statutory construction and constitutional expansions, the Supreme Court of India and various High Courts have laid down essential jurisprudence to govern digital likeness, identity theft, and algorithmic harms. This chapter analyzes the landmark judicial pronouncements that shape the contemporary regulatory horizon.

1. Constitutional Foundations of Digital Identity

Justice K.S. Puttaswamy (Retd.) v. Union of India (2017) 10 SCC 1

The foundational jurisprudence governing deepfakes stems from the landmark Right to Privacy

decision. The Supreme Court of India unanimously recognized privacy as an fundamental right under Article 21 of the Constitution.

Application to Deepfakes: The Court explicitly articulated three dimensions of privacy: spatial, informational, and bodily autonomy. Deepfakes fundamentally violate the core tenet of "informational privacy," which grants an individual the sovereign right to control their digital footprint and prevent the unauthorized algorithmic manipulation of their facial features and voice.

2. Personality Rights and Commercial Exploitation of AI

Anil Kapoor v. Simply Life India & Ors. (2023) FSR 12 (Del)

In this landmark decision, the Delhi High Court protected an individual's digital identity from unauthorized generative AI creation. The plaintiff sought a permanent injunction against several tech platforms that were using artificial intelligence to clone his voice, generate morphing pornography, and exploit his digital likeness without consent.

The Judicial Rationale: The Court held that an individual's name, voice, and unique persona cannot be stripped away or commercialized by third-party algorithms. The bench observed that while free speech allows creative parodies, it cannot cross the line into violating human dignity or executing commercial exploitation through non-consensual deep learning tools. This precedent establishes a strong tortuous remedy against digital identity theft.

3. Free Speech, Safe Harbor, and Platform Liability

Shreya Singhal v. Union of India (2015) 5 SCC 1

This case serves as a vital constitutional check on how the state regulates online content. While striking down Section 66A of the IT Act, the Supreme Court completely overhauled the doctrine of intermediary liability under Section 79.

Application to Deepfakes: The Court ruled that private social media platforms cannot exercise independent censorship or act as judges of truth. They are only obligated to take down content upon receiving an actual court order or a formal government directive. This judgment presents a direct structural conflict with the latest Intermediary Rules, which mandate that platforms

proactively remove "Synthetically Generated Information" (SGI) within narrow 2-to-3-hour windows, threatening over-censorship.

4. Criminality and Identity Theft Through Cyber Impersonation

SMC Pneumatics (India) Pvt. Ltd. v. Jogesh Kwatra (2001) Del Dist Ct

Though an early cyber law case, this decision established the Indian approach to digital impersonation and corporate identity fraud. The court penalized an employee who sent derogatory, fabricated corporate communications by creating spoofed digital entities.

Application to Deepfakes: The rationale laid down here emphasizes that digital impersonation intended to destroy a victim's socio-economic standing satisfies the threshold of criminal deception, serving as the historical baseline for prosecuting modern voice-cloning and deepfake fintech scams under Sections 66C and 66D of the IT Act

CRITICAL ANALYSIS AND DISCUSSION — THE DOCTRINAL GAP IN COGNIZING SYNTHETIC MALICE

The intersection of generative artificial intelligence and Indian penal statutes reveals a deep structural chasm between traditional criminal definitions and modern algorithmic harms. Indian courts are forced to rely on a fragmented patchwork of old laws to handle high-tech deepfakes. This critical analysis unpacks the core theoretical conflicts, enforcement hurdles, and structural flaws within India's current legal framework.

1. The Fallacy of Physicality and Tangible Harm

The most glaring conceptual vulnerability within the *Bharatiya Nyaya Sanhita (BNS), 2023* is its historical reliance on **physical metrics** and **tangible, material injuries**.

The Forgery Paradox: Under Section 336 of the BNS, the act of forgery requires the creation of a "false document or electronic record." However, generative adversarial networks (GANs) do not physically forge or copy an existing document. Instead, they ingest thousands of open-source data points to synthesize an entirely original array of pixels or audio frequencies.

The Intention Gap: Furthermore, traditional penal provisions for cheating by impersonation (Section 319, BNS) or identity theft (Section 66C, IT Act) are strongly tied to showing *mens*

rea for financial fraud, wrongful gain, or property loss. When a deepfake is deployed for non-financial harassment, political gaslighting, or digital defamation, the prosecution faces extreme difficulty satisfying this rigid statutory standard, leaving victims of reputational and psychological trauma without direct penal remedies.

2. The Misalignment of Gender Jurisprudence and Digital Realities

The weaponization of Non-Consensual Intimate Imagery (NCII) highlights how poorly traditional gender protection laws match automated digital harms.

Because the victim never actually performed the explicit act in the real world, defense lawyers argue that Section 77 of the BNS does not apply. This forces prosecutors to rely heavily on Sections 67 and 67A of the IT Act. However, these sections treat the issue as one of **public obscenity and morality** rather than recognizing it as a direct violation of a woman's **digital privacy, consent, and bodily dignity** under Article 21.

3. Executive Overreach and the Chilling Effect on Free Expression

To bypass the slow pace of the legislative process, the Ministry of Electronics and Information Technology (MeitY) has aggressively turned to executive rule-making under the *IT Intermediary Rules*. By labeling deepfakes as **Synthetically Generated Information (SGI)** and mandating aggressive 2-to-3-hour takedown windows, the state has effectively shifted the burden of truth onto private technology companies.

This model directly violates the core principles established by the Supreme Court in ***Shreya Singhal v. Union of India (2015)***. Faced with losing their statutory "safe harbor" immunity under Section 79 of the IT Act, platforms naturally resort to defensive, over-zealous automated filtering. This algorithmic censorship cannot understand the subtle differences of human communication, leading to the immediate suppression of legitimate political parodies, creative satire, and independent investigative journalism.

COMPREHENSIVE LEGISLATIVE AND STRUCTURAL SUGGESTIONS — REFORMING THE LEGAL INTERFACE

The preceding chapters reveal that India’s legal architecture is fundamentally ill-equipped to handle the challenges of generative artificial intelligence. Resolving the deepfake crisis cannot

be accomplished by modifying old laws or using aggressive platform regulations that threaten free speech under Article 19(1)(a). India requires a forward-looking, technology-sensitive statutory model.

This chapter outlines detailed, workable suggestions divided into three critical areas: substantive amendments to criminal law, procedural transformations in digital regulation, and the deployment of advanced cyber-forensic infrastructure.

1. Substantive Amendments to Criminal Law: Reforming the Bharatiya Nyaya Sanhita (BNS), 2023

The BNS must move away from its historical focus on physical proximity and tangible, material property loss to effectively penalize automated digital harms.

1.1 Codification of "Digital Identity and Persona Theft"

The legislature must introduce a new, dedicated provision—**Section 356A**—directly after the provision for criminal defamation. This section must explicitly criminalize the *unauthorized, non-consensual simulation, cloning, or distribution of a natural person's unique biometric voice frequency, facial features, or bodily likeness using digital algorithms*.

Decoupling Financial Fraud: The core of this offense must focus entirely on the violation of an individual's **digital dignity and informational privacy** under Article 21, completely removing the traditional requirement to prove financial loss or wrongful property gain.

1.2 Statutory Explanations for Generative Forgery

To stop defense lawyers from exploiting loopholes regarding "authorial authentication," an *Explanation* must be added to **Section 336 (Forgery)** and **Section 340 (Making a false electronic record)**. This explanation should state:

> *"Any individual who utilizes generative artificial intelligence, neural networks, or deep learning software to synthesize an electronic record that realistically mimics a living or deceased person's voice, image, or actions without their express consent, shall be deemed to have created a false electronic record with the intent to deceive."*

1.3 Redefining Sexual Harassment and Voyeurism for Synthetic Media

To protect victims of Non-Consensual Intimate Imagery (NCII), **Section 77 (Voyeurism)** must expand its statutory boundaries. The phrase "capturing or disseminating the image of a woman engaging in a private act" must be amended to include **"the synthetic generation or algorithmic mapping of a woman's face onto an explicit or obscene digital record without her consent."** This structural change moves the legal focus away from outdated obscenity metrics and places it where it belongs: on protecting individual consent and bodily autonomy.

2. Procedural and Digital Regulation: Balancing Takedowns with Free Speech

The current model administered by the Ministry of Electronics and Information Technology (MeitY) shifts an immense burden of truth onto private intermediaries. This approach must be completely overhauled to protect constitutional speech.

2.1 Establishing an Independent Content Review Board

Instead of forcing tech platforms to make rapid decisions on complex political or satirical content within 2 to 3 hours, India should establish an **Independent AI Content Review Board (CRB)**. This autonomous, quasi-judicial body—consisting of cyber-law experts, judicial officers, and digital forensics technicians—should serve as the primary authority for validating controversial takedown orders. This mechanism removes censorship power from private tech giants and aligns with the protections established in ***Shreya Singhal v. Union of India (2015)***.

2.2 Statutory Safe-Harbor for Parody, Satire, and Academic Expression

The regulatory definitions for **Synthetically Generated Information (SGI)** must include clear statutory protections for creative parodies, political satire, and bona fide research. If an uploader implements the required **10% visible/audible AI disclosure watermark**, intermediaries must be legally barred from executing automated, on-sight removals unless a formal order is issued by the Content Review Board.

3. Upstream AI Governance: Adopting Proactive, Product-Centric Frameworks

India must look beyond regulating downstream social media platforms and introduce strict legal obligations for the developer firms creating generative AI tools.

3.1 Mandating Technical Provenance Standards

The government must legally mandate that all generative AI platforms operating within Indian jurisdiction implement cryptographic metadata protocols, such as the **Coalition for Content Provenance and Authenticity (C2PA)** framework.

Every AI engine must automatically inject an unalterable, cryptographically signed digital fingerprint directly into the metadata of every generated file. This ensures that any piece of synthetic media can be instantly traced back to its software source, simplifying attribution during criminal investigations.

3.2 Graduated Risk Classification (The EU AI Act Hybrid Model)

India should enact a standalone **Digital Services and Artificial Intelligence Act** that adopts a graduated, risk-based classification model:

Risk Category	**Type of AI Technology**	**Statutory Obligation under Indian Law**
Unacceptable Risk	Automated, non-consensual deepfake pornography engines.	Absolute statutory ban and criminalization of development.
High Risk	Real-time voice cloning tools and conversational fintech bots.	Mandatory registration, source-code logging, and strict KYC verification.
Limited Risk	General entertainment filters and standard image generators.	Compulsory user disclosures and standard metadata watermarking.

4. Institutional and Evidence Law Reforms: Upgrading the Forensic Interface

4.1 Automated Forensic Infrastructure under the BSA, 2023

To effectively deploy the **Bharatiya Sakshya Adhiniyam, 2023 (BSA)**, India's law enforcement agencies must upgrade their technological infrastructure. The Ministry of Home

Affairs must set up a dedicated **National Deepfake Detection Cell (NDDC)** across all Central Forensic Science Laboratories (CFSL).

This cell must be equipped with automated deep-learning detection software capable of instantly checking for structural pixel irregularities and acoustic voice-cloning markers. This infrastructure will allow for the fast-tracked issuance of admissibility certificates under **Section 63 of the BSA**, matching the viral speed of digital crimes.

5. Conclusion of Suggestions

Implementing these comprehensive reforms will allow the Indian state to transition from a reactive, platform-dependent compliance regime to a proactive, rights-focused legal architecture. By modernizing definitions within the BNS, establishing a quasi-judicial content review board, and enforcing upstream metadata provenance tracking, India can robustly safeguard its citizens' fundamental rights under Article 21 without stifling technological innovation.

REFERENCE

I. Primary Sources: Legal Statutes, Rules & Bills (India)

The Bharatiya Nyaya Sanhita (BNS), 2023, No. 45 of 2023, §§ 77 (Voyeurism), 319 (Cheating by Impersonation), 336 (Forgery), 340 (False Electronic Record) (India).

The Bharatiya Sakshya Adhiniyam (BSA), 2023, No. 47 of 2023, § 63 (Admissibility of Electronic Records) (India).

The Information Technology Act, 2000, No. 21 of 2000, §§ 66C (Identity Theft), 66D (Cheating by Impersonation using Communication Resource), 66E (Privacy Violation), 67, 67A (Obscene/Sexually Explicit Material) (India).

The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (as amended), notified under Section 87, Information Technology Act, 2000 (Regulating Synthetically Generated Information [SGI] and Platform Takedown Deadlines).

The Digital Personal Data Protection (DPDP) Act, 2023, No. 26 of 2023 (India) (Regarding informational autonomy and non-consensual processing of personal biometric traits).

II. Judicial Precedents (Case Law)

Shreya Singhal v. Union of India, (2015) 5 SCC 1 (Striking down Section 66A of the IT Act and establishing strict boundaries for online free speech and intermediary safe harbor).

Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1 (Establishing the fundamental right to privacy, informational self-determination, and digital dignity under Article 21 of the Constitution of India).

III. Secondary Sources: Academic Journals & Law Reviews

Jorwal, Niharika. "Deepfake Technology and Criminal Law Reform in India: Addressing Synthetic Media Under the Bharatiya Nyaya Sanhita, 2023." *International Journal for Multidisciplinary Research (IJFMR)*, Vol. 8, Issue 2 (March–April 2026).

IJFMR *Focuses on the doctrinal ambiguities and need for a structured framework under the*

newly enforced BNS.

Singh, Preksha. "Deepfakes, identity theft, and the dark web: Legal gaps in AI-Generated fraud, an Indian perspective." *International Journal of Civil Law and Legal Research*, Vol. 5, Issue 2, Part B (2025): pp. 103–108.

Analyses legislative lacunae regarding AI voice-cloning, synthetic identities, and cyber-forensic shortfalls.

International Journal of Civil Law and Legal Research

Singh, Harmanjeet, & Ritu Panta. "Deepfake Evidence and the Indian Criminal Justice System: Challenges of Authenticity, Consent, and Admissibility in Law." *International Journal for Multidisciplinary Research (IJFMR)*, Vol. 7, Issue 6 (November 2025).

ResearchGate

Critically reviews evidentiary challenges of synthetic media under Section 63 of the Bharatiya Sakshya Adhiniyam (BSA), 2023.

Divyashree, N. R. "Deepfakes and Indian Criminal Law: Addressing the Gaps in Legal Protection." *Indian Journal of Law and Legal Research (IJLLR)*, Vol. 7, Issue 3 (2025).

Indian Journal of Law and Legal Research - IJLLR

Highlights the gaps in traditional Indian Penal Code/BNS provisions concerning non-consensual intimate imagery (NCII).

Record of Law

Editorial Board. "Criminal Liability for AI-Generated Deepfake Sexual Content under BNS: Gaps, Overlaps, and the Need for a Separate Offence." *Record of Law Review* (2025).

Advocates for a distinct statutory offense mapping the involuntary sexualization of identity to violations of Article 21.

IV. International Frameworks & Technical Standards

European Union AI Act: *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence*, OJ L 2024/1689.

Used for cross-border comparative analysis and the graduated, risk-based classification model recommended in your suggestions.

Coalition for Content Provenance and Authenticity (C2PA): *Technical Specifications for Asset and Metadata Provenance, v1.3* (2023).

The technical baseline referenced for watermarking and upstream cryptographic provenance tracking.