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

VOLUME 6 AND ISSUE 8 OF 2026

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

Published by

Institute of Legal Education

https://iledu.in

ADEQUACY OF INDIAN CYBER LAWS IN TACKLING DEEPFAKE CRIMES: A CRITICAL LEGAL ANALYSIS

AUTHOR – DHANALAKSHMI R* & AJAY KRISHNA. S.P**

* STUDENT AT VELS INSTITUTE OF SCIENCE, TECHNOLOGY & ADVANCED STUDIES (VISTAS)

** ASSISTANT PROFESSOR AT SCHOOL OF LAW, VELS INSTITUTE OF SCIENCE, TECHNOLOGY AND ADVANCED STUDIES (VISTAS)

BEST CITATION – DHANALAKSHMI R & AJAY KRISHNA. S.P, ADEQUACY OF INDIAN CYBER LAWS IN TACKLING DEEPFAKE CRIMES: A CRITICAL LEGAL ANALYSIS, *INDIAN JOURNAL OF LEGAL REVIEW (IJLR),* 6 (8) OF 2026, PG. 57-68, APIS – 3920 – 0001 & ISSN - 2583-2344.

ABSTRACT

The rapid evolution of Generative Artificial Intelligence has given rise to deepfakes—highly sophisticated synthetic media that can convincingly manipulate audio, video, and images to replicate real individuals with alarming realism. In India, deepfake technology has been weaponised for non-consensual pornography, financial fraud through voice cloning, electoral manipulation, and targeted harassment, generating urgent demands for legal intervention. This article critically evaluates the adequacy of Indian cyber law in addressing deepfake-related offences. Employing a doctrinal methodology, it analyses the Information Technology Act 2000, the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 and the Information Technology Amendment Rules 2026, the Bharatiya Nyaya Sanhita 2023, and the Digital Personal Data Protection Act 2023, situating these instruments within the constitutional framework of Articles 19(1)(a) and 21. The article identifies a critical 'detection-regulation gap,' wherein the law mandates rapid removal of harmful synthetic content without corresponding advancements in forensic detection capabilities, creating risks of over-censorship and disproportionate compliance burdens on smaller intermediaries. It further examines the evidentiary challenges inherent in establishing malicious intent and authenticating deepfake content under criminal law, and the incomplete criminological coverage of traditional offence categories—defamation, forgery, and impersonation—when applied to AI-generated synthetic media. Through comparative analysis with the European Union Artificial Intelligence Act, the article argues that India's framework remains predominantly reactive, addressing post-publication harm rather than pre-emptively regulating deepfake-enabling AI technologies. It concludes by proposing a multi-layered reform agenda: a dedicated Deepfake and Artificial Intelligence Regulation Act, adoption of a risk-based classification framework, establishment of a centralized AI Regulatory Authority, mandatory watermarking and content provenance standards, and a national deepfake forensic infrastructure.

Keywords: *Deepfakes, synthetic media, Information Technology Act 2000, IT Amendment Rules 2026, Synthetically Generated Information, Bharatiya Nyaya Sanhita 2023, Digital Personal Data Protection Act 2023, EU AI Act, detection-regulation gap, intermediary liability, Article 21, cyber law India*

I. INTRODUCTION

The digital age has witnessed an unprecedented transformation in the creation and dissemination of information, driven by rapid advances in Artificial Intelligence. Among these developments, deepfake technology has

INDIAN JOURNAL OF LEGAL REVIEW [IJLR – IF SCORE – 7.58]
VOLUME 6 AND ISSUE 8 OF 2026
APIS – 3920 - 0001 *(and)* ISSN - 2583-2344
Published by
Institute of Legal Education
https://iledu.in

emerged as one of the most disruptive and legally consequential innovations of our time. Deepfakes are AI-generated or algorithmically manipulated audio-visual representations that convincingly replicate real individuals—often indistinguishably from authentic media. While such technology holds legitimate applications in entertainment, education, and virtual communication, its misuse has generated profound legal, ethical, and constitutional concerns.

In India, the scale and variety of deepfake-related harm is alarming. Non-consensual deepfake pornography disproportionately victimises women; voice-cloning technologies have enabled sophisticated financial fraud by impersonating senior corporate officials; and manipulated political videos have been deployed to spread electoral misinformation on a mass scale.[1] The viral architecture of social media platforms amplifies these harms exponentially, rendering post-publication remedies increasingly inadequate.

The primary legislative framework governing cyber activities in India—the Information Technology Act 2000 (IT Act)—was enacted in a technological context that could not have anticipated the complexities of Generative AI and deepfake manipulation. Although the Act contains provisions addressing privacy violations, obscenity, identity theft, and intermediary liability, these provisions were designed for a fundamentally different digital landscape. Recognising this inadequacy, the Government of India introduced the Information Technology Amendment Rules 2026 (IT Rules 2026), which formally recognise 'Synthetically Generated Information' (SGI) and impose stringent obligations on intermediaries including mandatory labelling and accelerated takedown mechanisms.[2]

Yet critical problems persist. The requirement for content removal within two to three hours raises acute questions of technical feasibility, particularly for smaller intermediaries lacking sophisticated AI detection tools. The absence of a robust national forensic infrastructure to verify the authenticity of digital media creates a 'detection-regulation gap' in which legal mandates outpace technological capability. The Bharatiya Nyaya Sanhita 2023 (BNS)—which replaced the Indian Penal Code 1860—provides traditional criminal offences that can be extended to deepfake conduct through judicial interpretation, but does not explicitly address AI-generated crime. The Digital Personal Data Protection Act 2023 (DPDP Act) provides complementary data-protection safeguards but does not directly criminalise deepfake creation.[3]

This article undertakes a critical legal analysis of the adequacy of Indian cyber law in addressing deepfake crimes. Part II examines the conceptual and technical dimensions of deepfakes and the constitutional framework within which regulation must operate. Part III analyses the pre-2026 legal regime and its structural limitations. Part IV evaluates the paradigm shift introduced by the IT Rules 2026, with particular attention to the SGI framework and its implementation challenges. Part V assesses the criminal law dimension under the BNS 2023 and the DPDP Act, including evidentiary challenges. Part VI undertakes a comparative analysis with the European Union AI Act. Part VII proposes a comprehensive reform agenda.

II. THE CONCEPTUAL AND CONSTITUTIONAL FRAMEWORK

A. Understanding Deepfakes: Technology and Typology

The term 'deepfake' is a portmanteau of 'deep learning' and 'fake,' signifying AI-generated content produced through neural network techniques—principally Generative Adversarial Networks (GANs), autoencoders, and, more recently, diffusion models. GANs operate through two competing neural networks: a generator that produces synthetic content and a discriminator that evaluates its authenticity; through iterative competition, the generator

produces outputs of progressively greater realism.[4]

Deepfakes may be classified into several principal categories. Face-swapping deepfakes superimpose one individual's facial features onto another's body in video or image content—the dominant form employed in non-consensual pornography and political misinformation. Voice-cloning deepfakes replicate an individual's speech patterns, tone, and accent with high fidelity, and have been used in corporate fraud and financial impersonation scams. Lip-syncing deepfakes alter existing video footage to synchronise with a fabricated audio track, fabricating statements never made by the depicted individual. Full-body deepfakes manipulate the movements and gestures of an entire person. AI-generated text—while not traditionally classified as deepfakes—frequently complements audio-visual synthetic media to create comprehensive misinformation ecosystems.[5]

These technologies are becoming increasingly accessible. Open-source software, cloud-based AI platforms, and mobile applications now enable individuals with limited technical expertise to generate convincing synthetic media, dramatically lowering the barrier to misuse and increasing the scale of potential harm.

B. The Constitutional Framework for Deepfake Regulation

Deepfake regulation in India operates at the intersection of two fundamental constitutional values: the right to privacy and dignity under Article 21, and the right to freedom of speech and expression under Article 19(1)(a). Both values constrain and shape the permissible scope of legislative intervention.

In Justice K.S. Puttaswamy (Retd.) v. Union of India,[6] a nine-judge bench of the Supreme Court unanimously recognised the right to privacy as a fundamental right under Article 21, encompassing personal autonomy, informational privacy, and dignity. The Court's recognition of 'decisional autonomy' as a protected interest—the right of individuals to control their own identity and personal information—has direct implications for deepfake regulation. Deepfakes violate decisional autonomy by fabricating the identity, voice, and likeness of real individuals without consent, subjecting them to reputational harm and psychological trauma that strikes at the core of the right to live with dignity.

Article 19(1)(a) simultaneously constrains regulatory overreach. While the Constitution permits restrictions on speech on grounds enumerated in Article 19(2)—including public order, decency, and morality—such restrictions must be reasonable and proportionate. Mandatory takedown regimes, mandatory labelling, and intermediary liability rules all engage Article 19(1)(a) and must withstand the proportionality standard articulated by the Supreme Court in Shreya Singhal v. Union of India.[7] The risk of over-censorship—where legitimate expression, satire, parody, and journalistic comment are suppressed alongside genuinely harmful deepfakes—is a structural danger that any regulatory framework must actively guard against.

III. THE PRE-2026 LEGAL FRAMEWORK AND ITS STRUCTURAL LIMITATIONS

A. Overview of the Information Technology Act 2000

The IT Act was enacted to provide legal recognition to electronic records and digital signatures, facilitate e-commerce, and penalise cyber offences. Amended significantly in 2008, it established a framework for punishing hacking, identity theft, and publication of obscene material in electronic form. However, its conceptual architecture was built around the early internet—a period of static web pages, email, and basic transactional platforms—rather than the dynamic, AI-powered content ecosystem of today.[8]

The Act's offence provisions relevant to deepfakes operate indirectly. Section 66E

penalises the violation of privacy by capturing and transmitting images of private areas without consent—applicable to deepfake pornography only where identifiable private-area imagery is reproduced, but not to wholly fabricated synthetic content. Section 66C (identity theft) and Section 66D (cheating by personation) may be invoked where voice cloning or face-swapping is used for fraudulent impersonation, but proof of 'dishonest intention' faces significant evidentiary challenges in deepfake contexts, where the creator may be anonymous and geographically distant. Sections 67, 67A, and 67B penalise obscene and sexually explicit electronic content and child pornography, and have been applied to deepfake pornography, but their reach is limited to obscenity-threshold content and does not extend to political misinformation, reputational deepfakes, or financial fraud.[9]

B. Intermediary Liability and the Safe Harbour Regime

Section 79 of the IT Act provides conditional immunity to intermediaries—social media platforms, search engines, and hosting providers—from third-party content hosted on their platforms, provided they observe due diligence, do not initiate the transmission, and comply with government guidelines. The Supreme Court in Shreya Singhal clarified that intermediaries are obligated to act upon receipt of a court order or notification from a government authority, not upon private complaint alone—a principle designed to prevent arbitrary censorship.[10]

This framework proved structurally inadequate in the deepfake context. Deepfake content spreads at viral speed; by the time a court order is obtained or government notification issued, irreversible harm to reputation, privacy, or democratic processes may already have occurred. The safe harbour's reactive character—acting after content is already circulating—rather than requiring proactive detection and removal—reflects a regulatory philosophy unsuited to AI-generated synthetic media capable of causing instantaneous mass harm.

C. Critical Assessment of the Pre-2026 Regime

The pre-2026 legal framework suffers from three structural deficiencies in addressing deepfakes. First, legislative lacuna: the complete absence of any statutory definition or recognition of deepfakes, synthetic media, or AI-generated content leaves enforcement agencies and courts without a defined legal category, compelling reliance on indirect and often strained interpretations of existing offence provisions. Second, reactive enforcement: the framework operates exclusively ex post, addressing harm after dissemination rather than preventing creation or enabling rapid intervention at the point of spread. Third, evidentiary inadequacy: the attribution of deepfake authorship, the authentication of AI-generated content, and the establishment of criminal intent all require forensic capabilities—metadata analysis, AI detection tools, chain-of-custody protocols for digital evidence—that the Act neither mandates nor facilitates.[11]

IV. THE PARADIGM SHIFT: THE INFORMATION TECHNOLOGY AMENDMENT RULES 2026

A. The Concept of Synthetically Generated Information

The IT Rules 2026 introduce the concept of 'Synthetically Generated Information' (SGI) as the foundational legal category for regulating AI-generated and AI-manipulated content. SGI is defined to include any audio, video, image, or text content created or substantially altered using artificial intelligence, machine learning, or other automated digital technologies in a manner that makes it appear authentic or human-generated.[12]

The legal recognition of SGI represents the first instance in Indian law of a statutory category specifically directed at AI-generated digital harm. By creating this distinct classification, the 2026 framework enables targeted regulatory interventions—mandatory labelling, accelerated takedown, and enhanced intermediary

obligations—specifically tailored to the characteristics of synthetic media, rather than relying on the interpretive extension of offence provisions designed for entirely different forms of conduct. The introduction of SGI provides the conceptual architecture upon which a more comprehensive deepfake regulatory regime can eventually be constructed.

B. Mandatory Labelling Requirements

The Rules require all intermediaries handling SGI to ensure that AI-generated or AI-manipulated content is clearly and visibly identified before it is made accessible to users, whether through visible disclaimers, metadata tagging, or embedded digital watermarks.[13] This labelling obligation reflects the principle of transparency in digital communication—users must be in a position to make informed judgments about the content they consume.

The mandatory labelling requirement imposes a significant technological obligation on platforms: they must deploy or procure AI-based detection systems capable of identifying synthetic media with sufficient accuracy to enable consistent labelling compliance. This is technically demanding, given that state-of-the-art deepfake detection systems—while improving rapidly—remain imperfect, producing both false positives (legitimate content labelled as synthetic) and false negatives (synthetic content escaping detection). False positives threaten freedom of expression by wrongly stigmatising authentic content; false negatives defeat the regulatory purpose entirely.[14] A labelling regime premised on imperfect detection technology risks systematic failures on both sides of this binary.

C. Accelerated Takedown Mechanisms

The 2026 Rules mandate time-bound removal of unlawful or injurious SGI upon receipt of a valid complaint from a user, affected individual, or government authority. For sensitive categories of content—content threatening individual dignity or involving sexually explicit deepfakes—removal timelines as short as two to three hours are prescribed.[15] This represents a fundamental departure from the reactive, court-order-dependent removal regime of the pre-2026 framework, reflecting a legislative recognition that the viral dynamics of digital content render delayed enforcement functionally equivalent to no enforcement at all.

However, the accelerated takedown mechanism raises acute concerns of proportionality and over-censorship. Intermediaries facing severe liability for non-compliance with two-to-three-hour timelines will rationally choose to remove content pre-emptively in borderline cases—removing content whose legality is uncertain rather than risking penalties. This 'risk-averse compliance' dynamic suppresses legitimate speech alongside harmful content, producing a chilling effect on expression that is constitutionally problematic. The Supreme Court's articulation in Shreya Singhal of the necessity for judicial or governmental authorisation before takedown—designed precisely to prevent private censorship by intermediaries—sits in tension with a regime that incentivises rapid private removal without such authorisation.

D. Compliance Burden on Intermediaries and the Detection-Regulation Gap

The 2026 Rules significantly expand intermediary obligations: deploying advanced SGI detection systems, appointing compliance officers, maintaining detailed content moderation logs, and establishing real-time grievance redressal infrastructure. While these requirements enhance accountability, they create disproportionate compliance burdens for small and medium-sized digital platforms that lack the technological and financial resources of large technology corporations.[16] The result may be a regulatory landscape dominated by large incumbents, with smaller platforms either exiting the market or abandoning safe harbour protection—an outcome antithetical to the goal of promoting a diverse and innovative digital ecosystem.

INDIAN JOURNAL OF LEGAL REVIEW [IJLR – IF SCORE – 7.58]
VOLUME 6 AND ISSUE 8 OF 2026
APIS – 3920 - 0001 *(and)* ISSN - 2583-2344
Published by
Institute of Legal Education
https://iledu.in

The most fundamental structural problem is the detection-regulation gap: the law mandates rapid removal of SGI content but does not mandate, fund, or facilitate the development of the forensic detection infrastructure needed to implement that removal reliably and accurately. India lacks a centralised, standardised national deepfake forensic infrastructure. Without such infrastructure, intermediaries are required to rely on privately procured AI detection tools of varying accuracy, creating systemic inconsistency in enforcement and the risk of both systematic over-censorship and systematic under-enforcement.

V. THE CRIMINAL LAW DIMENSION: BNS 2023, THE DPDP ACT, AND ARTICLE 21

A. Deepfake Offences under the Bharatiya Nyaya Sanhita 2023

The Bharatiya Nyaya Sanhita 2023 replaced the Indian Penal Code 1860 with the objective of modernising India's criminal law. While the BNS adopts a largely technology-neutral approach that enables its provisions to be applied to digitally committed offences, it contains no provisions explicitly addressing deepfakes or AI-generated synthetic media.[17] The criminalisation of deepfake conduct therefore proceeds through interpretive extension of traditional offence categories.

Defamation under the BNS is particularly relevant where deepfakes are deployed to fabricate videos or audio recordings depicting an individual engaging in illegal, immoral, or socially damaging conduct. The widespread dissemination of such content on social media platforms significantly aggravates the harm relative to traditional defamation, as the viral architecture of digital platforms multiplies the audience for the fabricated portrayal and the highly realistic nature of deepfakes increases audience credulity. However, criminal defamation requires proof of the accused's intent to harm reputation, and the anonymity of online deepfake creators—frequently operating through pseudonymous accounts and encrypted platforms—makes establishing the accused's identity and intent exceptionally difficult.

Forgery provisions under the BNS may be applied to deepfakes that constitute fabricated electronic records intended to deceive—for example, a manipulated video of an official statement used to facilitate fraud or political manipulation. Impersonation offences are applicable where deepfake technology is used to replicate an individual's face, voice, or mannerisms for fraudulent purposes. Courts are required to extend interpretations designed for physical or identity-based impersonation to the entirely novel domain of AI-driven digital identity replication.[18]

The critical limitation of the BNS framework is its indirect and interpretive character. The absence of an explicit offence of 'deepfake creation' or 'malicious synthetic media dissemination' means that the applicability of specific offences to specific deepfake conduct must be determined case by case, creating legal uncertainty, inconsistency in enforcement, and reduced deterrent effect. The elements of traditional offences—particularly mens rea requirements—were designed for entirely different factual contexts and impose evidentiary burdens that are structurally misaligned with the technical realities of deepfake production.

B. Evidentiary Challenges in Deepfake Prosecutions

The evidentiary challenges in prosecuting deepfake offences are formidable and constitute perhaps the most significant practical obstacle to effective criminal enforcement. The prosecution bears the burden of proving beyond reasonable doubt not only that the content is artificially generated, but also that the accused created or disseminated it with the requisite criminal intent.[19]

Establishing the authenticity or inauthenticity of deepfake content requires advanced forensic techniques: analysis of metadata embedded in digital files, examination of pixel-level

inconsistencies produced by AI generation algorithms, voice pattern analysis in audio deepfakes, and application of AI-based detection tools that identify tell-tale artefacts of GAN or diffusion model generation. These techniques are probabilistic rather than determinative, producing confidence-weighted assessments rather than binary conclusions. The reliability of AI-generated forensic findings as evidence in criminal proceedings—and the admissibility requirements under the Bharatiya Sakshya Adhiniyam 2023—present unresolved questions of evidentiary law.[20]

Attribution—establishing that the accused created the deepfake—is frequently defeated by the combination of pseudonymous online accounts, virtual private networks, encrypted communications platforms, and cross-border hosting of content on servers outside Indian jurisdiction. The collection of evidence from foreign servers requires invocation of mutual legal assistance treaty mechanisms, which are notoriously slow and often ineffective in addressing the time-sensitive demands of digital criminal investigation.

C. The Digital Personal Data Protection Act 2023 and Personality Rights

The DPDP Act 2023 establishes a framework for the lawful processing of personal data, requiring that personal data be processed only for lawful purposes and with valid consent except in specified circumstances. The unauthorised use of an individual's photographs, video recordings, voice recordings, or biometric data to generate deepfakes constitutes unlawful processing of personal data under the Act—a direct violation of the data principal's right to informational self-determination.[21] The Act provides rights of correction, erasure, and grievance redressal, enabling victims of deepfake-based data misuse to seek administrative remedies.

However, the DPDP Act operates as a civil and regulatory mechanism rather than a penal statute. It does not criminalise deepfake creation or dissemination, and its enforcement proceeds through Data Protection Board proceedings rather than criminal prosecution. Its contribution to combating deepfake crimes is therefore preventive and complementary—limiting access to the personal data that deepfakes exploit—rather than directly punitive.

The doctrine of personality rights, progressively developed through judicial interpretation of Article 21, provides a further constitutional foundation for legal protection against identity manipulation through deepfakes. Following the Supreme Court's expansive interpretation of Article 21 in Puttaswamy, courts have recognised rights to dignity, reputation, and control over personal identity as components of the right to life and personal liberty. Deepfakes directly violate these rights by enabling the fabrication and mass dissemination of false representations of real individuals. In the absence of dedicated statutory protection, personality rights claims provide an important—if judicially uncertain—avenue for injunctive and compensatory relief.[22]

VI. INTERNATIONAL PERSPECTIVES: THE EU AI ACT AND LESSONS FOR INDIA

A. Overview of the EU Artificial Intelligence Act

The European Union Artificial Intelligence Act (EU AI Act), adopted in 2024, constitutes the world's first comprehensive, binding AI-specific regulatory framework. It adopts a risk-based classification model that categorises AI systems into four tiers: systems posing unacceptable risks (which are prohibited outright), high-risk systems (subject to strict compliance obligations), limited-risk systems (subject primarily to transparency obligations), and minimal-risk systems (largely unregulated). Deep-fake technology, and generative AI systems capable of producing synthetic content depicting real individuals, fall within the limited-risk and, in certain applications, high-risk categories.[23]

The EU AI Act's approach to synthetic media is architecturally preventive: it imposes transparency and disclosure obligations at the

design and deployment stage, before AI systems reach users. Providers of AI systems capable of generating synthetic audio, video, or imagery of individuals must ensure that such content is technically marked—through watermarking or digital provenance metadata—and that users are clearly informed of its synthetic nature. This ex-ante regulatory model contrasts sharply with India's predominantly ex-post framework, which intervenes after harmful content has been created and disseminated rather than at the point of AI system design.

B. Comparative Analysis with the Indian Framework

The contrast between the EU AI Act and the Indian legal framework is striking across four dimensions. First, regulatory unity: the EU Act provides a single, unified regulatory instrument for AI governance, with clear hierarchical obligations, centralized enforcement authority, and harmonized implementation across member states. India's framework is fragmented across five or more instruments—the IT Act 2000, the IT Rules 2026, the BNS 2023, the DPDP Act 2023, and constitutional provisions—without a single coordinating regulatory authority, resulting in overlapping jurisdiction and enforcement gaps.[24]

Second, temporal orientation: the EU model is predominantly ex-ante—preventive compliance before deployment—while India's model is predominantly ex-post—reactive enforcement after harm. The IT Rules 2026 introduce faster takedown mechanisms, but these operate within the same reactive paradigm, accelerating the response to harm rather than preventing it. Third, technological specificity: the EU Act directly regulates deepfake-enabling AI systems at the design and deployment stage, requiring technical safeguards such as watermarking to be built into the technology. India regulates the content produced by such systems without directly regulating the AI systems themselves, leaving the upstream source of harm unaddressed. Fourth, institutional infrastructure: the EU Act is backed by centralized, well-resourced regulatory authorities with technical expertise. India lacks a dedicated AI regulatory authority, distributing enforcement responsibilities across courts, cybercrime cells, CERT-In, and intermediaries.

C. Challenges in Adopting the EU Model and Lessons for India

Direct transplantation of the EU AI Act framework to India is neither feasible nor desirable. India's vastly larger and more heterogeneous digital user base, its diverse linguistic and socio-economic landscape, the limited technical capacity of smaller intermediaries, and the constitutional constraints imposed by Articles 19(1)(a) and 21 all require a context-specific regulatory approach.[25] Compliance costs that are manageable for large European technology companies may be prohibitive for Indian startups and small digital platforms. A rigid, compliance-heavy framework risks suppressing technological innovation and reinforcing market concentration.

The EU experience nonetheless offers four crucial lessons for India. First, legal integration: the superiority of a unified AI regulatory statute over a fragmented multi-statute approach. Second, ex-ante regulation: the importance of regulating AI systems capable of generating synthetic media at the design and deployment stage rather than exclusively at the content dissemination stage. Third, technical mandates: the value of directly mandating technical safeguards—watermarking, digital provenance—as non-negotiable features of deepfake-capable AI systems. Fourth, institutional centralisation: the necessity of a dedicated AI regulatory authority to coordinate enforcement, develop technical standards, and ensure consistent implementation across platforms and sectors.

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

VOLUME 6 AND ISSUE 8 OF 2026

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

Published by

Institute of Legal Education

https://iledu.in

VII. CRITICAL ANALYSIS AND PROPOSED REFORMS

A. Enacting a Dedicated Deepfake and AI Regulation Act

The most urgent legislative reform is the enactment of a standalone Deepfake and Artificial Intelligence Regulation Act that provides clear statutory definitions of 'deepfake' and 'synthetic media,' establishes specific offences for the creation, manipulation, and malicious dissemination of deepfakes with graded penalties calibrated to harm severity, directly regulates AI systems capable of generating synthetic media at the design and deployment stage, and confers exclusive jurisdiction on a specialist court or tribunal for deepfake-related offences. A dedicated statute would eliminate interpretive ambiguity, ensure prosecutorial clarity, and establish a coherent enforcement framework that the current multi-statute approach cannot provide.

B. Adopting a Risk-Based Classification Framework

India should adopt a risk-based AI classification framework analogous to the EU AI Act, categorising deepfake-capable AI applications by their potential for harm. High-risk applications—AI systems used for electoral manipulation, financial impersonation fraud, non-consensual sexual content, and biometric identity replication—should be subject to mandatory pre-deployment registration, risk assessment, and technical certification. Low-risk applications—AI tools for entertainment, artistic expression, education, and research with appropriate safeguards—should be subject to lighter regulatory obligations to preserve innovation and creative freedom. This calibrated approach enables proportionate regulation without the blunt-instrument risks of a uniform prescriptive framework.

C. Establishing a Centralised AI Regulatory Authority

A dedicated Artificial Intelligence Regulatory Authority should be established with statutory powers to monitor AI systems across sectors, issue binding technical standards for deepfake detection and content provenance, coordinate enforcement among law enforcement agencies, CERT-In, and the Data Protection Board, maintain a national register of high-risk AI applications, and adjudicate intermediary compliance disputes. Centralising AI governance in a single expert body would eliminate the fragmentation and coordination failures that characterise the current multi-agency framework. The Authority should be staffed by personnel with technical AI expertise and legal qualifications, ensuring that regulatory decisions are grounded in technological understanding as well as legal principle.

D. Building National Deepfake Forensic Infrastructure

The detection-regulation gap—the mismatch between legal mandates for rapid content removal and available forensic detection capabilities—can only be addressed through sustained investment in national deepfake forensic infrastructure. This should include: establishment of certified national deepfake detection laboratories with standardised protocols for authenticating digital content; development of national technical standards for digital evidence in AI-related criminal proceedings, addressing chain of custody, detection accuracy thresholds, and the admissibility of AI-generated forensic findings; mandatory technical training for law enforcement cyber cells, public prosecutors, and judicial officers in AI forensics and digital evidence evaluation; and government-funded research into deepfake detection technologies, particularly methods robust to the rapidly evolving sophistication of AI generation models.

E. Mandatory Watermarking and Content Provenance

All AI systems deployed in India that are capable of generating synthetic audio, video, or imagery of real individuals should be required to implement industry-standard digital

watermarking and content provenance mechanisms as non-negotiable technical features. Content Credentials—open provenance standards maintained by the Coalition for Content Provenance and Authenticity (C2PA)—provide a technically mature framework for embedding unforgeable metadata into AI-generated content, enabling reliable identification and attribution. Mandating such standards at the AI system level—rather than relying on post-hoc labelling by intermediaries—would address the detection-regulation gap at its source and ensure that synthetic content carries verifiable provenance regardless of the platform on which it is subsequently disseminated.

F. Constitutional Safeguards and the Proportionality Principle

Any reformed deepfake regulatory framework must rigorously protect freedom of expression under Article 19(1)(a). The following safeguards are essential: explicit exemptions for satire, parody, artistic expression, journalistic comment, and academic research, with a clearly defined 'public interest' defence; a proportionality review requirement, ensuring that all regulatory interventions are necessary, suitable, and no more restrictive of speech than required to achieve the legitimate regulatory objective; independent judicial or quasi-judicial oversight of takedown orders, preventing the transformation of intermediaries into private censors; and a structured appeals mechanism for wrongful content removal with remedies including reinstatement, compensation, and costs.

VIII. FINDINGS AND CONCLUSION

A. Summary of Findings

This article has established the following principal findings. First, the pre-2026 Indian cyber law framework—anchored in the IT Act 2000—is structurally inadequate to address deepfake crimes, suffering from legislative lacuna (no statutory recognition of synthetic media), reactive enforcement (post-dissemination remedies only), and evidentiary inadequacy (forensic infrastructure insufficient to support criminal prosecution). Second, the IT Rules 2026 represent a significant regulatory advance through the introduction of the SGI framework, mandatory labelling, and accelerated takedown mechanisms, but are undermined by the detection-regulation gap, risks of over-censorship, and disproportionate compliance burdens on smaller intermediaries. Third, the BNS 2023 provides traditional criminal offences that can be extended to deepfake conduct through judicial interpretation, but their indirect and interpretive application—combined with formidable evidentiary challenges—produces legal uncertainty and reduced deterrent effect. Fourth, the DPDP Act 2023 provides complementary preventive safeguards through data protection principles but does not directly criminalise deepfake creation. Fifth, the comparative analysis with the EU AI Act demonstrates the superiority of unified, ex-ante, risk-based AI regulation and exposes the structural fragmentation of India's current approach. Sixth, a comprehensive reform agenda—dedicated legislation, risk-based classification, centralised regulatory authority, national forensic infrastructure, mandatory watermarking, and constitutional safeguards—is both necessary and achievable.

B. Conclusion

The adequacy of Indian cyber law in tackling deepfake crimes must be assessed honestly: the current framework is partially adequate at best, and fundamentally reactive at worst. The IT Rules 2026 demonstrate genuine regulatory ambition and represent a meaningful step forward from the legislative vacuum of the pre-2026 regime. But ambition and adequacy are not synonymous. A framework that mandates two-to-three-hour content removal without providing reliable detection infrastructure, that criminalises deepfake harm through provisions designed for pre-digital offences, and that lacks a unified regulatory instrument or dedicated enforcement authority is a framework whose

architecture is misaligned with the technology it seeks to govern.

The deepfake challenge is not merely technical. It is fundamentally a question of constitutional values: how the State can protect the right to privacy and dignity guaranteed by Article 21 without suppressing the right to free expression guaranteed by Article 19(1)(a); how criminal law can establish liability for novel AI-enabled harms without abandoning the evidentiary standards that protect the innocent; and how India can regulate at the necessary scale and speed without building regulatory structures that entrench existing power asymmetries between large and small actors in the digital economy.

The reform agenda proposed in this article—a dedicated Deepfake and AI Regulation Act, a risk-based classification framework, a centralised AI Regulatory Authority, national forensic infrastructure, mandatory content provenance, and robust constitutional safeguards—provides a comprehensive and constitutionally grounded framework for meeting these challenges. India's digital ecosystem is too large, too dynamic, and too consequential for its governance to remain in permanent catch-up mode with the technologies it seeks to regulate. The time for a proactive, integrated, and technologically informed legal response is now.

BIBLIOGRAPHY

A. Primary Sources — Statutes and Constitutional Provisions

Constitution of India 1950, Articles 19(1)(a), 19(2), 21.

Information Technology Act 2000 (No 21 of 2000), ss 66C, 66D, 66E, 67, 67A, 67B, 79.

Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021.

Information Technology (Amendment) Rules 2026.

Bharatiya Nyaya Sanhita 2023 (No 45 of 2023).

Bharatiya Sakshya Adhiniyam 2023 (No 47 of 2023).

Digital Personal Data Protection Act 2023 (No 22 of 2023).

European Union Artificial Intelligence Act (Regulation (EU) 2024/1689).

General Data Protection Regulation (Regulation (EU) 2016/679).

B. Cases

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

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

Avnish Bajaj v. State (NCT of Delhi) (2005) 116 DLT 427.

Subramanian Swamy v. Union of India (2016) 7 SCC 221.

C. Books

Duggal P, Cyber Law: The Indian Perspective (Saakshar Law Publications 2023).

Kumar A, Law Relating to Cyber Crimes and Electronic Evidence (LexisNexis 2022).

Myneni SR, Cyber Law (Asia Law House 2021).

Sharma V, Information Technology Law and Practice (Universal Law Publishing 2020).

Farid H, Photo Forensics (MIT Press 2019).

Zuboff S, The Age of Surveillance Capitalism (PublicAffairs 2019).

Lessig L, Code and Other Laws of Cyberspace (Basic Books 1999).

D. Journal Articles and Reports

Internet Freedom Foundation, 'Understanding Deepfakes in India' (IFF Report 2024).

NITI Aayog, Discussion Paper on Artificial Intelligence for India (NITI Aayog 2023).

OECD, Principles on Artificial Intelligence (OECD 2023).

World Economic Forum, Global Risks Report 2024 (WEF 2024).

Partnership on AI, Guidelines for Synthetic Media (PAI 2023).

Law Commission of India, Report on Technology and Law (Law Commission 2017).

Ministry of Electronics and Information Technology, Consultation Paper on AI Governance (MeitY 2024).

Gillespie T, 'The Politics of Platforms' (2012) 14(4) New Media & Society 347.

E. Websites and Online Sources

www.meity.gov.in | www.indiacode.nic.in | www.scconline.com

www.manupatra.com | www.livelaw.in | www.indiankanoon.org

www.unodc.org | www.oecd.org/digital/artificial-intelligence

ENDNOTES

1 Internet Freedom Foundation, 'Understanding Deepfakes in India' (IFF Report 2024) 3–7.

2 Information Technology (Amendment) Rules 2026, Rule 3(1)(b)(ix).

3 Digital Personal Data Protection Act 2023 (No 22 of 2023).

4 Hany Farid, 'Detecting Digital Forgeries' (MIT Technical Report 2019) 2–5.

5 World Economic Forum, Global Risks Report 2024 (WEF 2024) 14.

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

7 Shreya Singhal v. Union of India (2015) 5 SCC 1, [81]–[93].

8 Vakul Sharma, Information Technology Law and Practice (Universal Law Publishing 2020) 45 48.

9 Information Technology Act 2000, ss 66C, 66D, 66E, 67, 67A, 67B.

10 Shreya Singhal v. Union of India (2015) 5 SCC 1, [117]–[120].

11 Law Commission of India, Report on Technology and Law (Law Commission 2017) [3.4]–[3.9].

12 Information Technology (Amendment) Rules 2026, Rule 2(1)(sa).

13 ibid Rule 3(1)(b)(ix).

14 Hany Farid, 'Photo Forensics' (MIT Press 2019) 112–115.

15 Information Technology (Amendment) Rules 2026, Rule 3(2)(b).

16 Ministry of Electronics and Information Technology, Consultation Paper on AI Governance (MeitY 2024) [4.2].

17 Bharatiya Nyaya Sanhita 2023 (No 45 of 2023).

18 Pavan Duggal, Cyber Law: The Indian Perspective (Saakshar Law Publications 2023) 134–138.

19 Arvind Kumar, Law Relating to Cyber Crimes and Electronic Evidence (LexisNexis 2022) 213.

20 Bharatiya Sakshya Adhiniyam 2023 (No 47 of 2023), s 63.

21 Digital Personal Data Protection Act 2023 (No 22 of 2023), s 4.

22 Justice K.S. Puttaswamy (Retd.) v. Union of India (2017) 10 SCC 1, [301]–[303] (Chandrachud J).

23 European Union Artificial Intelligence Act (Regulation (EU) 2024/1689), Arts 5, 6, 50.

24 NITI Aayog, Discussion Paper on Artificial Intelligence for India (NITI Aayog 2023) 28–31.

25 OECD Principles on Artificial Intelligence (OECD 2023) Principle 1.5.