DEEPFAKES AND THE LAW: ANALYSING THE LEGAL VACUUM AND THENEED FOR REGULATORY REFORM IN INDIA

Manas Pratim Talukdar, BA LLB (Hons.), NERIM Law College

ABSTRACT

The rapid advancement of artificial intelligence has given rise to a deeply concerning technological phenomenon deepfakes. Deepfakes are hyper-realistic synthetic media, typically video or audio, generated by machine learning systems that superimpose a person's likeness upon another body or fabricate statements attributed to real individuals. While the technology carries legitimate applications in entertainment and education, its misuse poses grave threats to individual dignity, privacy, democratic integrity, and national security. India, as one of the world's largest digital economies, faces acute vulnerability to deepfake abuse, yet its legal framework remains inadequate to address this challenge. This essay examines the nature and harms of deepfakes; critically analyses existing Indian legal provisions under the Information Technology Act 2000, the Bharatiya Nyaya Sanhita 2023, and the Digital Personal Data Protection Act 2023; identifies significant lacunae therein; develops a conceptual argument grounding deepfake harm within the doctrine of "representational autonomy" under Article 21; applies the constitutional proportionality test to the proposed regulatory response; draws critically evaluated lessons from the United States, United Kingdom, and European Union; and ultimately advocates for a dedicated, comprehensive Deepfake Regulation Act for India, stress-tested against enforcement feasibility.

Keywords: Deepfakes, Artificial Intelligence, Cyber Law, Representational Autonomy, IT Act, Privacy, Regulatory Reform, India.

I. INTRODUCTION

The digital age has witnessed the emergence of technologies that blur the boundary between authentic reality and engineered fabrication. Among the most alarming of these is the deepfake a portmanteau of "deep learning" and "fake" which employs neural network architectures to generate synthetic audio-visual content virtually indistinguishable from genuine material. [1] The legal significance of this technology lies not in its mechanics but in its consequences: it enables the fabrication of consent, the annihilation of reputation, and the manufacture of political reality, at scale and with minimal technical expertise.

Scholars have warned of a "liar's dividend" a paradox in which the very existence of deepfake technology allows malicious actors to dismiss genuine evidence as fabricated, thereby destabilising evidentiary frameworks that legal systems depend upon. In India, deepfake abuse has already manifested: a 2023 incident involving a morphed video of actress Rashmika Mandanna went viral before detection, triggering public outrage. [2] The government's response an advisory from the Ministry of Electronics and Information Technology was reactive and non-binding.[3] This gap between technological threat and legal response is the central concern of this essay.

This essay argues that India's existing legal architecture, though not entirely silent on fraud, privacy, and defamation, is structurally incapable of addressing the unique, multi-dimensional, and real-time harms posed by deepfakes. The argument proceeds through doctrinal analysis, constitutional theory, comparative law, and feasibility assessment, converging on the case for dedicated legislation.

II. UNDERSTANDING DEEPFAKES: NATURE, TYPOLOGY, AND HARM

Deepfakes can be categorised into three principal types: face-swap deepfakes, which digitally superimpose a person's face onto another's body; audio deepfakes, which clone a voice to fabricate statements never made; and full-body deepfakes, involving comprehensive manipulation of movement and expression. Each carries distinct but overlapping legal

[1]Robert Chesney & Danielle Keats Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security" (2019) 107 California Law Review 1753, 1758.
[2]Press Trust of India, "Deepfake videos of actresses trigger debate on regulation", The Hindu (November 2023).
[3]Ministry of Electronics and Information Technology, Advisory on Deepfakes (November 2023), Government of India.

implications.

The most prevalent and damaging application has been non-consensual intimate imagery (NCII) commonly termed "deepfake pornography." Studies indicate that approximately 96% of all deepfake videos online constitute non-consensual pornographic content, with women bearing the overwhelming burden of this abuse. [4] Beyond sexual exploitation, deepfakes have been weaponised for political disinformation, financial fraud through voice-cloned "CEO fraud," evidence fabrication, and targeted harassment of journalists and activists. The distinguishing characteristic of deepfake harm relative to conventional defamation or fraud is its virality, irreversibility, and the psychological impact of seeing one's own likeness weaponised against oneself.

III. REPRESENTATIONAL AUTONOMY: A CONCEPTUAL FRAMEWORK

Before proceeding to doctrinal analysis, this essay advances a conceptual argument that deepfakes constitute a violation of what may be termed "representational autonomy" the right of an individual to exercise sovereign control over their digital likeness and the narratives attached to it. This concept emerges from, but extends beyond, the existing doctrines of informational privacy and dignitary harm.

The Supreme Court in Puttaswamy recognised that privacy encompasses not merely the right to be left alone, but the right to control the dissemination of personal information and to protect one's identity from appropriation. [5] Justice Kaul, concurring, articulated a right to "determine for oneself when, how and to what extent information about oneself is communicated to others." [6] Representational autonomy situates deepfakes squarely within this framework: when a deepfake manufactures a person's voice, face, or body performing acts they did not perform or speaking words they did not utter, it does not merely misrepresent it confiscates. The victim's digital identity is expropriated and weaponised without consent.

This framing carries concrete doctrinal implications. First, it grounds deepfake harm in Article 21 rather than merely in tort, elevating the constitutional imperative for legislative intervention. Second, it distinguishes deepfakes from conventional defamation: defamation involves false

[4]Sensity AI, "The State of Deepfakes 2023: Landscape, Threats, and Impact" (2023) 4.
[5]Justice K.S. Puttaswamy (Retd.) v Union of India (2017) 10 SCC 1, [180] per Chandrachud J.
[6]ibid [310] per Kaul J, articulating the right to control the dissemination of personal information.

statements about a person; a deepfake involves false embodiment of a person. The harm is not simply reputational but existential it violates the victim's capacity to be the author of their own public identity. Third, it provides a principled basis for extending protection beyond intimate imagery to any unauthorised use of a person's likeness whether for political manipulation, commercial exploitation, or identity fraud. Regulators who frame deepfake legislation around representational autonomy will produce more coherent and constitutionally durable law than those who treat the problem as merely a subset of obscenity or fraud.

IV. THE EXISTING LEGAL FRAMEWORK IN INDIA: A CRITICAL ASSESSMENT

India has no legislation specifically addressing deepfakes. Several existing statutes have been tentatively applied to deepfake-related conduct, but each reveals structural limitations when tested against the unique features of synthetic media.

A. The Information Technology Act, 2000

Section 66C penalises identity theft involving electronic records, [7] while Section 66E criminalises the violation of privacy through the non-consensual capture or transmission of images of a person's private areas. [8] Sections 67 and 67A penalise the publication of obscene and sexually explicit electronic material. [9] These provisions, however, were drafted in 2000, before the advent of generative AI, and their textual limitations are severe. Section 66E requires the "capture" of an actual private area a condition that synthetic deepfake content, which generates rather than captures, may not technically satisfy. Section 66C requires dishonest or fraudulent use of a genuine electronic signature or password; it does not contemplate the manufacture of a synthetic digital identity. The Act's silences are not mere gaps they are structural incompatibilities with the deepfake problem.

B. The Bharatiya Nyaya Sanhita, 2023

The BNS contains provisions on cheating (Section 318) and defamation (Section 356) that may be invoked in deepfake cases involving financial fraud or reputational damage respectively. [10] However, these are generic provisions requiring proof of intent and harm through conventional

[7]Information Technology Act 2000, s 66C (India).
[8]Information Technology Act 2000, s 66E (India).
[9]Information Technology Act 2000, ss 67, 67A (India).
[10]Bharatiya Nyaya Sanhita 2023, ss 318, 356 (India).

evidentiary frameworks. They impose significant burdens on victims particularly in cases of viral deepfake content and offer no mechanism for emergency injunctive relief. The BNS also introduces no concept of synthetic or AI-generated evidence, leaving courts without statutory guidance on how to assess the authenticity of digital exhibits.

C. The Digital Personal Data Protection Act, 2023

The DPDPA 2023 establishes a framework requiring consent for personal data processing and imposing obligations on "data fiduciaries." [11] Biometric data used to train deepfake models could, in principle, constitute "personal data" under the Act. Yet the DPDPA addresses data collection not content generation or distribution. An actor who trains a model on unlawfully obtained biometric data violates the Act; one who generates a deepfake from publicly available images does not clearly do so. This gap renders the DPDPA an incomplete instrument against deepfake production.

D. Constitutional Dimensions and the Proportionality Test

The constitutional foundation for deepfake regulation lies in the Puttaswamy judgment, which established privacy including informational privacy and the right to control one's likeness as a fundamental right under Article 21. Any regulatory framework must, however, simultaneously withstand challenge under Article 19(1)(a), which protects freedom of speech and expression. Satire, political commentary, parody, and artistic expression may employ altered or synthetic imagery of public figures. The question, therefore, is not whether to regulate, but how to do so without unconstitutionally chilling protected expression.

Applying the Proportionality Test to Deepfake Regulation

The Supreme Court in Puttaswamy articulated a four-part proportionality test for evaluating restrictions on fundamental rights: (i) legitimate aim; (ii) rational nexus; (iii) necessity; and (iv) balancing. [12] A proposed Deepfake Regulation Act survives this test as follows.

First, legitimate aim is clearly established: the protection of individual dignity, privacy, and the

[11]Digital Personal Data Protection Act 2023 (India).
[12]Puttaswamy (n 11) [310]; the four-part proportionality test requires: (i) a legitimate aim; (ii) a rational nexus between the measure and the aim; (iii) necessity — the measure must be the least restrictive means; and (iv) proportionality stricto sensu — a fair balance between the right infringed and the aim pursued.

integrity of democratic discourse constitute compelling state interests of constitutional magnitude, recognised both in Puttaswamy and in the directive principles under Article 51A(e) and (f).

Second, rational nexus is self-evident: criminalising the non-consensual creation and distribution of deepfake content directly addresses the identified harm. Mandatory platform disclosure obligations and watermarking requirements are instrumentally connected to the aim of detection and accountability.

Third, necessity requires that the chosen measure be the least restrictive means of achieving the aim. This is where the drafting of any deepfake law is most demanding. A blanket prohibition on synthetic media would clearly fail this prong. The proposed Act must be carefully tailored: it should criminalise only non-consensual or deceptive use of synthetic media, with explicit carve-outs for satire, parody, artistic expression, and academic or journalistic use. The UK's Online Safety Act 2023 provides a useful model it does not prohibit deepfakes, but prohibits their non-consensual sharing.[13]

Fourth, balancing under Article 19(2) is satisfied when the restriction is proportionate to the harm. Given that deepfake-enabled NCII causes severe, often irreversible harm to victims psychological trauma, professional ruin, and the erasure of digital identity the proportionality of targeted regulation is well-established. The Supreme Court's jurisprudence in Anuradha Bhasin v Union of India confirms that digital restrictions must be narrowly tailored but are constitutionally permissible where they serve compelling interests. [14] A well-drafted Deepfake Regulation Act satisfies this standard.

V. KEY LEGAL LACUNAE

The critical gaps in India's existing framework are as follows. First, there is no statutory definition of "deepfake" or "synthetic media," leaving courts and enforcement agencies without definitional clarity. Second, no specific offence addresses the creation, distribution, or commission-with-intent of deepfake content prosecutions must be forced through ill-fitting provisions. Third, no civil cause of action specific to deepfake victims exists, leaving them

[13]Online Safety Act 2023 (UK), s 66; the Act creates a specific offence of sharing intimate deepfake imagery without consent, removing the requirement to prove intent to cause distress.
[14]Anuradha Bhasin v Union of India (2020) 3 SCC 637 (India), applying necessity and proportionality to digital restrictions.

dependent on defamation suits that impose high evidentiary burdens and offer no emergency relief. [15] Fourth, intermediary obligations under the IT Rules 2021 are too broadly worded to mandate proactive deepfake detection. [16] Fifth, India's law is entirely silent on jurisdictional complexity deepfakes may be created in one country, hosted in another, and consumed in a third. Sixth, there is no mechanism for rapid, court-supervised takedown of viral deepfake content; the window of maximum harm is often measured in hours, not the weeks or months that civil litigation requires.

VI. INTERNATIONAL REGULATORY MODELS: A CRITICAL EVALUATION

A survey of international approaches reveals meaningful progress alongside significant enforcement failures both of which are instructive for India.

A. United States

The United States has adopted a fragmented, state-by-state approach. The federal Deepfake Accountability Act proposed mandatory watermarking of synthetic media, while the DEFIANCE Act 2024 created a federal civil cause of action for victims of non-consensual intimate deepfakes. [17] The American model's principal weakness is jurisdictional incoherence: without uniform federal criminal law, enforcement varies dramatically across states, and platforms operating nationally face regulatory arbitrage. The experience also demonstrates that civil remedies alone are insufficient deterrents criminal sanctions are necessary to address the most severe violations. For India, the lesson is that piecemeal, sectoral amendment is less effective than a unified statute with both criminal and civil dimensions.

B. United Kingdom

The Online Safety Act 2023's criminalisation of non-consensual intimate deepfake sharing without requiring proof of intent to cause distress represents a victim-centric advance. However, the Act addresses distribution rather than creation, leaving open the question of liability for those who generate but do not share harmful deepfakes. Its enforcement is also

[15]Law Commission of India, Report No. 270 on "Hate Speech" (2017) 54 — noting enforcement asymmetries in technology-mediated offences.
[16]Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021, Rule 3(1)(b) (India).
[17]Deepfake Accountability Act, H.R. 4065, 116th Cong. (2019) (USA); DEFIANCE Act 2024, Pub. L. No. 118-73 (USA).

constrained by Ofcom's regulatory capacity and the global nature of platform hosting. For India, the UK model is valuable for its intent-free standard on distribution, but must be supplemented with provisions targeting creation and with mandatory takedown timelines applicable to platforms.

C. European Union

The EU's Artificial Intelligence Act 2024 imposes transparency obligations, including mandatory labelling of AI-generated content intended for public consumption. [18] The Act's risk-based architecture calibrating obligations to the potential severity of harm is conceptually sophisticated. Its principal limitation, however, is the assumption of a well-resourced compliance infrastructure that many jurisdictions, including India, do not yet possess. The Act also relies on voluntary corporate disclosure in many contexts, which experience suggests is insufficient without strong enforcement mechanisms. For India, the EU model provides a useful architectural template, but must be adapted to an enforcement environment characterised by under-resourced regulators, a vast informal digital economy, and widespread use of open-source AI tools that fall outside traditional regulatory perimeters.

VII. RISKS OF OVERCRIMINALISATION AND CHILLING EFFECT

Serious engagement with the deepfake problem requires honest acknowledgment of the risks that poorly drafted regulation would create. Three concerns are particularly significant.

First, there is a risk of chilling satire, parody, and artistic expression. A significant tradition of political commentary from cartoon animation to theatrical impersonation involves the alteration or mimicry of public figures' appearances and voices. Deepfake regulation that fails to carve out these forms of protected expression would unconstitutionally restrict political speech. The response to this concern is legislative precision: the proposed Act must define its prohibited conduct by reference to non-consent and deceptive intent, with explicit safe harbours for clearly labelled satire, parody, and transformative artistic works. The standard should be objective whether a reasonable person would understand the content as synthetic commentary, not as purporting to represent reality.

[18]European Parliament, Artificial Intelligence Act (2024), Regulation (EU) 2024/1689, Arts 50, 52 (transparency obligations on deepfake disclosure).

Second, overcriminalisation poses risks of misuse: vaguely drafted offences can be weaponised by powerful actors including the state against journalists, critics, or political opponents who produce legitimate synthetic media. [19] This risk demands that any deepfake legislation include a robust proportionality requirement and strong judicial oversight of criminal enforcement, rather than broad prosecutorial discretion.

Third, the evidential challenge of proving that content is synthetic rather than genuine — creates a risk of wrongful prosecution. Courts and enforcement agencies will require institutional capacity to engage technical expert evidence on deepfake detection. These concerns do not counsel against legislation — they counsel for careful, precise, and constitutionally conscious drafting.[20]

VIII. RECOMMENDATIONS FOR REFORM

The foregoing analysis converges on six recommendations, presented here in integrated form.

First, India must enact a dedicated Deepfake Regulation Act. This statute should provide a clear statutory definition of deepfake and synthetic media; establish specific criminal offences for the non-consensual creation and distribution of harmful deepfake content, with graduated penalties calibrated to harm severity; create a statutory civil cause of action for victims, with a presumption in favour of the claimant where non-consent is established; and include explicit safe harbours for satire, parody, artistic expression, and journalism, conditioned on clear labelling. The justification for a standalone Act rather than amendment of existing statutes is doctrinal coherence: deepfake harm involves a unique intersection of identity, consent, technology, and distribution that existing provisions cannot accommodate without significant judicial creativity.

Second, intermediary obligations must be substantially enhanced. The IT Rules 2021 should be amended to require large platforms to deploy AI-based deepfake detection tools, establish expedited takedown mechanisms with mandatory response windows of 24 hours for intimate content, provide clear disclosure labels on AI-generated material, and submit periodic transparency reports on deepfake removal. Proportionate compliance timelines longer for

[19]Danielle Keats Citron & Mary Anne Franks, "Criminalizing Revenge Porn" (2014) 49 Wake Forest Law Review 345, 381.
[20]Lilian Edwards, "Revenge Porn: The Criminal Justice and Courts Act 2015 and Beyond" (2015) 3 Journal of Media Law 1, 14 — noting the limited deterrent effect of criminal sanctions absent civil remedies.

smaller platforms would address concerns about compliance burdens on emerging digital businesses.

Third, a mandatory provenance and watermarking framework should require all AI platforms operating in India to embed verifiable, tamper-resistant metadata into AI-generated content. This would enable both users and enforcement agencies to trace the origin of synthetic media and establish chains of accountability. The EU's AI Act provides a working model for such a requirement.

Fourth, a Deepfake Adjudication Authority or dedicated fast-track courts with technical expertise should be established to provide rapid relief to victims. This body should have power to grant emergency injunctions against platforms within hours of a verified complaint, order compensation, and hear appeals against platform takedown decisions. The existing civil litigation infrastructure is poorly suited to the real-time dynamics of viral deepfake harm.

Fifth, India's criminal procedure must be updated to address the authentication of digital evidence in deepfake cases. The Indian Evidence Act and its successor, the Bharatiya Sakshya Adhiniyam 2023 should be amended to provide statutory standards for the forensic authentication of video and audio evidence, and to establish the admissibility of AI-generated forensic analysis as expert testimony.

Sixth, public digital literacy must be treated as a legal priority. The government should mandate media literacy programmes in educational institutions, equipping citizens to identify synthetic media and understand their legal rights. Awareness is both a preventive measure and a condition of effective legal enforcement: detection of deepfakes by ordinary users is currently a critical early-warning mechanism in the absence of automated platform detection at scale.

IX. IMPLEMENTATION FEASIBILITY: STRESS-TESTING THE PROPOSALS

Any honest reform agenda must acknowledge the enforcement challenges that India's socio-legal context presents.

The most significant structural challenge is the proliferation of open-source AI models. Unlike proprietary AI platforms subject to corporate compliance regimes, open-source tools freely available on decentralised repositories can be downloaded and run locally by any user, entirely

outside the regulatory perimeter of intermediary obligations.[21] This renders platform-centric regulation partially effective at best. The proposed Act must therefore supplement platform obligations with individual criminal liability for non-consensual creation, regardless of the tool used.

A second challenge is encrypted and anonymous distribution. Deepfake content routinely circulates through end-to-end encrypted messaging applications and anonymous social media accounts, making attribution extremely difficult. The proposed adjudication authority must be empowered to issue disclosure orders to platforms requiring the identification of account holders in verified deepfake cases, subject to judicial supervision. A calibrated tracing mechanism limited to criminal investigations and victim claims would balance enforcement effectiveness against privacy concerns.

A third challenge is jurisdictional. Deepfake content is frequently created abroad, hosted on foreign servers, and distributed through platforms incorporated outside India. The Deepfake Regulation Act must therefore include provisions for extra-territorial application criminalising conduct outside India where the victim is an Indian resident and establish bilateral cooperation frameworks with foreign jurisdictions. India's existing Mutual Legal Assistance Treaty network provides a foundation, but technology-specific protocols will be necessary.

A fourth challenge concerns regulatory capacity. Effective enforcement of the proposed Act will require significant investment in the technical capacity of law enforcement agencies and the judiciary. Courts adjudicating deepfake disputes will require access to accredited forensic experts. The proposed Deepfake Adjudication Authority must be staffed by individuals with both legal and technical expertise. Building this capacity is a medium-term project that must begin immediately, in anticipation of legislation.

X. CONCLUSION

Deepfakes constitute a category of harm that existing legal doctrine was not designed to address. They violate what this essay has termed representational autonomy the constitutionally grounded right to sovereign control over one's digital identity and likeness. They do so at a scale, speed, and degree of realism that renders conventional remedies in defamation, privacy,

[21]Oxford Internet Institute, "Open-Source AI and the Accountability Gap" (2023) Working Paper — documenting the proliferation of unregulated AI tools on decentralised platforms.

and fraud inadequate. The absence of a dedicated Indian legal framework is, therefore, not merely a technical gap to be patched by judicial creativity or executive advisory; it is a structural failure that leaves an expanding category of fundamental rights violations without a legal remedy calibrated to the nature of the harm.

The proportionality analysis conducted in this essay demonstrates that targeted deepfake regulation is constitutionally permissible under Articles 19 and 21, provided it is precisely drafted, includes robust safe harbours for protected expression, and is subject to judicial oversight. The comparative analysis reveals that no existing international model is wholesale transplantable to India's socio-legal context, but that the UK's victim-centric standard, the EU's risk-based architecture, and the American experience with civil remedies together provide a rich template for Indian legislators to adapt critically rather than copy.

The proposed Deepfake Regulation Act supported by enhanced intermediary obligations, a provenance framework, a dedicated adjudication mechanism, and investment in enforcement capacity represents a coherent, constitutionally sound, and practically feasible response. The case for legislation is not merely utilitarian. It is grounded in the proposition that a democratic state committed to constitutional values cannot permit the systematic weaponisation of a citizen's own identity against them, without remedy, at the scale that deepfake technology makes possible.