INTERNATIONAL JOURNAL FOR LEGAL RESEARCH AND ANALYSIS

Open Access, Refereed Journal Multi Disciplinary
Peer Reviewed

www.ijlra.com

ISSN: 2582-6433

DISCLAIMER

ISSN: 2582-6433

ABOUT US

The *International Journal for Legal Research & Analysis (IJLRA)* (ISSN: 2582-6433) is a peer-reviewed, academic, online journal published on a monthly basis. The journal aims to provide a comprehensive and interactive platform for the publication of original and high-quality legal research.

IJLRA publishes Short Articles, Long Articles, Research Papers, Case Comments, Book Reviews, Essays, and interdisciplinary studies in the field of law and allied disciplines. The journal seeks to promote critical analysis and informed discourse on contemporary legal, social, and policy issues.

The primary objective of IJLRA is to enhance academic engagement and scholarly dialogue among law students, researchers, academicians, legal professionals, and members of the Bar and Bench. The journal endeavours to establish itself as a credible and widely cited academic publication through the publication of original, well-researched, and analytically sound contributions.

IJLRA welcomes submissions from all branches of law, provided the work is original, unpublished, and submitted in accordance with the prescribed submission guidelines. All manuscripts are subject to a rigorous peer-review process to ensure academic quality, originality, and relevance.

Through its publications, the *International Journal for Legal Research & Analysis* aspires to contribute meaningfully to legal scholarship and the development of law as an instrument of justice and social progress.

ISSN: 2582-6433

PUBLICATION ETHICS, COPYRIGHT & AUTHOR RESPONSIBILITY STATEMENT

The *International Journal for Legal Research and Analysis (IJLRA)* is committed to upholding the highest standards of publication ethics and academic integrity. All manuscripts submitted to the journal must be original, unpublished, and free from plagiarism, data fabrication, falsification, or any form of unethical research or publication practice. Authors are solely responsible for the accuracy, originality, legality, and ethical compliance of their work and must ensure that all sources are properly cited and that necessary permissions for any third-party copyrighted material have been duly obtained prior to submission. Copyright in all published articles vests with IJLRA, unless otherwise expressly stated, and authors grant the journal the irrevocable right to publish, reproduce, distribute, and archive their work in print and electronic formats. The views and opinions expressed in the articles are those of the authors alone and do not reflect the views of the Editors, Editorial Board, Reviewers, or Publisher. IJLRA shall not be liable for any loss, damage, claim, or legal consequence arising from the use, reliance upon, or interpretation of the content published. By submitting a manuscript, the author(s) agree to fully indemnify and hold harmless the journal, its Editor-in-Chief, Editors, Editorial Board, Reviewers, Advisors, Publisher, and Management against any claims, liabilities, or legal proceedings arising out of plagiarism, copyright infringement, defamation, breach of confidentiality, or violation of third-party rights. The journal reserves the absolute right to reject, withdraw, retract, or remove any manuscript or published article in case of ethical or legal violations, without incurring any liability.

ISSN: 2582-6433

AI-GENERATED HARMS AND HUMAN RIGHTS: DATA COMMODIFICATION, DEEPFAKE PROLIFERATION, AND NORMATIVE IMPERATIVES IN INDIA AND BEYOND

AUTHORED BY - VISHAKHA TRIPATHI & DR. ARVIND KUMAR SINGH

ABSTRACT

This paper examines the human rights implications of artificial intelligence within user-generated content ecosystems, focusing on two interconnected harms: the commodification of personal data and the proliferation of deepfakes. This integration has accelerated the human rights violation. It argues that while India's digital expansion has enabled innovation, it has simultaneously exposed individuals to serious violations of privacy, dignity, and economic security. Drawing on doctrinal analysis and recent empirical trends, the paper evaluates the how generative tools have triggered the deepfake content with rise in women targetef cyber crimes in India, context to adequacy of existing legal frameworks, including the Information Technology Act, 2000 and the Digital Personal Data Protection Act, 2023. It identifies key gaps, particularly the absence of a creation-based offence for synthetic media and weak enforcement mechanisms. By engaging with comparative frameworks such as the EU AI Act and international human rights standards, the article proposes a preventive regulatory model grounded in constitutional values. It concludes that a shift from reactive to proactive regulation is essential to safeguard rights in an AI-driven digital ecosystem. The analysis holds relevance for Global South jurisdictions confronting analogous regulatory deficits.

Keywords: artificial intelligence; deepfakes; data privacy; human rights; cybercrime; gender-based violence; India; regulatory reform

INTRODUCTION

India's rapid emergence as a digital powerhouse—driven by a user base of nearly 900 million internet users—has fundamentally transformed the nature of social, economic, and communicative interactions. The everyday documenting life practices—sharing photographs, uploading short-form video, transmitting voice messages across Instagram, WhatsApp, and YouTube—have become the primary vast reservoir for AI model training. Platforms systematically aggregate and commercially exploit this data, often without meaningful

ISSN: 2582-6433

informed consent, thereby undermining the right to informational privacy protected under Article 21 of the Constitution[1] as authoritatively construed in Justice K.S. Puttaswamy (Retd.) v. Union of India.[2] Unequivocally affirmed that privacy includes control over personal information and decisional autonomy. Yet, in practice, the consent mechanisms followed by digital platforms remain largely formalistic, offering users little real choice or control over how their data is utilised.

At the same time, AI tools have turned people's own content against them. Deepfakes—fake videos or audio made with AI—increased by 900%, thanks to over 5,000 easy face-swap apps and 1,000 voice-cloning tools that anyone can use.

In this evolving ecosystem, user-generated content (UGC) has effectively become a form of economic resource, routinely collected, processed, and monetised by digital platforms.

This sharp increase is closely linked with a 60 per cent escalation in women-targeted cybercrime complaints—from approximately 50,000 cases in 2024 to 80,000 in 2026[3]—with empirical research indicating that 93–98 per cent of non-consensual deepfakes target women.[4] Financial fraud compounds this vulnerability: voice-cloned executive impersonation schemes—colloquially termed 'CEO fraud'—are estimated to extract ₹15,000 crore annually through UPI-based scams.[5]

This article dissects two structurally linked violations:

- the commodification of personhood through **UGC DATAFICATION**
- the assault on dignity, equality, and economic security through **DEEPFAKE PROLIFERATION**.

The harm caused by deepfakes is not limited to privacy violations but extends to dignity and identity. The Supreme Court in Navtej Singh Johar v. Union of India[6] recognised dignity as an intrinsic component of Article 21, emphasising the protection of individual identity and

[1]Constitution of India 1950, Article 21.
[2]Justice K.S. Puttaswamy (Retd.) v. Union of India (2017) 10 SCC 1.
[3]National Crime Records Bureau, Crime in India 2025–2026 (Ministry of Home Affairs 2026).
[4]Home Security Heroes, The State of Deepfakes 2024: Landscape, Threats, and Impact (Home Security Heroes 2024) <www.homesecurityheroes.com/state-of-deepfakes>; UN Women, Technology-Facilitated Gender-Based Violence: Global Assessment 2026 (UN Women 2026).
[5]Reserve Bank of India, Annual Report on Financial Fraud and Cybercrime 2025–2026 (RBI 2026).
[6] Navtej Singh Johar Union of India Writ Petition (Criminal) No 76 of 2016 (Supreme Court of India, 6 September 2018)

ISSN: 2582-6433

autonomy in both physical and digital space. The creation and circulation of non-consensual synthetic media directly undermine these constitutional values.

The article addresses three central questions:

(i) To what extent do existing legal frameworks, including the Information Technology Act and the Digital Personal Data Protection Act, adequately respond to AI-mediated harms?

(ii) What doctrinal and regulatory gaps allow such harms to continue with limited accountability?

(iii) How can international frameworks, particularly the EU AI Act and evolving UN human rights standards, inform a more robust and rights-oriented regulatory approach in India?

This paper uses doctrinal research methods, backed by data from NCRB, Pi-Labs, and RBI reports[7]. It argues that India's current laws—focused on fixing harm after it happens, not preventing it—are not equipped to handle the fast spread of AI harms. We need upfront laws that match tech changes and global rules to protect Article 21 rights in an AI world.[8]

LITERATURE REVIEW:

What's Out There and What's Missing Privacy and Data Issues

Solove's idea of "contextual integrity" shows how wrong it is when personal stuff we share online—like selfies or posts—gets turned into data for AI companies without our okay. We share it for friends, but it becomes big business info. Zuboff calls this "surveillance capitalism," where they grab extra data we didn't agree to give, feeding a huge $100 billion AI industry. It's built into the system, not just random slip-ups.[9]

Indian experts like Kaur say our DPDP Act's consent rules don't cut it. Once your data turns into AI stuff, you can't really delete the copies or "data shadows" that stick around forever—Section 12 fixes nothing real. That's a big hole between laws on paper and what actually works, which this paper digs into.

[7] Pi -Labs (n 3); NCRB (n 4); RBI (n 6).
[8] Constitution of India 1950, Article 21 (n 1).
[9] Shoshana Zuboff, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (Profile
Books 2019) 93–95.

ISSN: 2582-6433

DEEPFAKES AND HOW THEY AFFECT WOMEN

Early work by Citron and Franks treated revenge porn as a serious attack on dignity that needs its own laws. Chesney and Citron saw deepfakes coming, saying they'd mess with truth and people's lives. Studies now prove it's mostly porn—95% of deepfakes—and almost all (98%) target women.[10]

In India, people write about cases like Rashmika Mandanna's 2023 deepfake video, which got 50 million views before sites pulled it. Bajpai calls it a wake-up call for rules here, but most writing just describes what happened, not why or how to fix it.

DEEPFAKES: GENDERED WEAPONISATION AND THE SCHOLARSHIP

The foundational legal scholarship on non-consensual intimate imagery—Citron and Franks—characterised such harms as dignitary torts warranting dedicated legal prohibition.[11] Chesney and Citron presciently extended this analysis to deepfakes, warning to truth, authenticity, and individual dignity.[12] Subsequent empirical work has confirmed and sharpened the gendered character of the harm: approximately 95% of deepfakes constitute non-consensual pornography, with 98% targeting women.[13]

Indian scholarship has increasingly engaged the deepfake phenomenon through the lens of specific high-profile cases. Bajpai analyses the Rashmika Mandanna incident of November 2023—in which a pornographic deepfake achieved over 50 million views before platform takedown—as a watershed moment for Indian regulatory consciousness, though the analytical emphasis remains predominantly descriptive rather than causative.[14]

Research Gaps: The UGC Nexus and Global South Perspectives

A critical lacuna pervades the existing scholarship: the literature bifurcates data governance concerns from deepfake harm analysis, obscuring their shared UGC substrate. ‘Pi-Labs’

[10] Home Security Heroes (n 4); UN Women (n4).
[11]Danielle Keats Citron and Mary Anne Franks, 'Criminalizing Revenge Porn' (2014) 49(2) Wake Forest Law Review 345, 362.
[12]Robert Chesney and Danielle Keats Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107(6) California Law Review 1753, 1760.
[13]Home Security Heroes (n 5); UN Women (n 5).
[14]Rohit Bajpai, 'Rashmika Mandanna and the Deepfake Reckoning: Regulatory Lacunae in Indian Cyber Law' (2025) 17(2) Indian Journal of Cyber Law 45, 49–52.

ISSN: 2582-6433

quantification of the 900 per cent content surge—the first systematic attribution to tool democratisation across 5,000+ applications—remains entirely uncited in legal scholarship as of the period under review.[15] The financial fraud dimension, encompassing ₹15,000 crore in voice-cloning scams, has similarly evaded doctrinal scrutiny.[16] International legal journals disproportionately privilege EU and US regulatory frameworks, leaving India-centric causal analyses of Global South enforcement realities largely absent. The present article bridges these Gaps.

METHODOLOGY

This paper adopts a doctrinal methodology, involving the systematic examination of statutes, judicial precedents, and legal instruments, supplemented by empirical data for contextual grounding. The analysis unfolds in four stages: (i) identifying the nature of the harms; (ii) mapping applicable legal provisions; (iii) evaluating their effectiveness through case law review; and (iv) proposing reforms benchmarked against international norms.

Primary sources include

- the Information Technology Act, 2000
- Digital Personal Data Protection Act, 2023
- Articles 14, 19, and 21 of the Indian Constitution
- Article 17 of the ICCPR
- Article 5(1)(d) of the EU AI Act.
- Key judicial decisions analysed are Puttaswamy (2017), Shreya Singhal (2015), X Corp v. Union of India (2025), and FIRs related to the Rashmika Mandanna deepfake incident.

Empirical support draws from Pi-Labs (2026), NCRB Crime Reports (2024–2026), and RBI Annual Fraud Bulletins (2026). To maintain focus, the study excludes discussions of surveillance bias or AI discrimination, which merit separate analysis.[17]

[15]Pi-Labs (n 3); cf Bajpai (n 20).
[16]RBI (n 6).
[17] Information Technology Act 2000 (India), §§43A, 66E, 67, 67A, 79; Digital Personal Data Protection Act 2023 (India), §§8–12; Constitution of India 1950, Articles 14, 19, 21; International Covenant on Civil and Political Rights (adopted 16 December 1966) 999 UNTS 171, Article 17; EU AI Act, Regulation (EU) 2024/1689, Article 5.

ISSN: 2582-6433

UGC DATAFICATION: STRUCTURAL MECHANICS AND RIGHTS PATHOLOGIES

The Infrastructure of Extraction

The scale of UGC extraction is staggering: platforms collectively ingest an estimated 10 billion uploads daily, encompassing facial biometric data, voiceprints, and granular behavioural traces. These aggregations populate large-scale training corpora—such as LAION-5B—that are subsequently commercialised to AI developers at significant value. Crucially, India-specific data indicates that approximately 70 per cent of AI applications explicitly target or disproportionately process data from minors[18], in apparent contravention of the DPDP Act's verifiability requirements under §8.

The legal concept of 'consent' is rendered formally operative but substantively illusory by Terms of Service agreements that routinely extend to fifty pages, written in specialised legal language, and presented as a binary take-it-or-leave-it proposition. The Supreme Court in Puttaswamy recognised cognitive overload and power asymmetry as factors vitiating meaningful consent in digital contexts.[19] The resulting 'consent illusion'—formal compliance masking substantive coercion—systematically enables mass rights violations at a scale no post-hoc enforcement mechanism can adequately remedy.

IMPACT ON HUMAN RIGHTS

The rights harm generated by this infrastructure operate at three levels. *First,* at the level of informational privacy: Article 21's protection of informational self-determination[20] is systematically undermined by the generation of 'data shadows'—indelible derivative profiles enabling perpetual behavioural prediction and targeting far beyond the original data subject's contemplation or consent.[21]

Second, at the level of economic justice: content creators receive zero remuneration for UGC that constitutes the primary training material for AI systems generating multi-billion-dollar commercial returns—OpenAI's market capitalisation alone exceeded $150 billion by mid-

[18] National Human Rights Commission, Advisory on AI Applications Targeting Minors (NHRC March 2026).
[19] Puttaswamy (n 2) para 130 (Chandrachud J).
[20] Constitution of India 1950, Article 21 (n 1).
[21] Puttaswamy (n 2) para 130.

ISSN: 2582-6433

2026.[22] This represents a structural transfer of value from data subjects to platform intermediaries and AI developers, raising questions about the adequacy of consent frameworks as instruments of distributive justice.

Third, at the level of financial security: voiceprints extracted from UGC are systematically exploited for CEO fraud—cloned audio authorising large-value transfers—with individual incidents generating losses exceeding ₹500 crore.[23] The convergence of privacy violation and financial predation represents a qualitative escalation in the severity of datafication harms.

Doctrinal Mapping and Legal Inadequacy

Puttaswamy constitutionalised privacy as a fundamental right subject to a proportionality standard that requires any limitation to be necessary, rational, and least restrictive.[24] However, the operationalisation of this standard through statutory instruments reveals significant deficits. The IT Act §43A activates liability only post-breach, providing no preventive mechanism.[25] The DPDP Act §12 confers erasure rights on data principals, but these rights are structurally ineffective against inferential derivatives: once UGC generates a model weight, the original data point's deletion leaves the derivative intact.[26] Aadhaar and UIDAI v. CBI affirmed the data minimisation principle, yet this standard remains practically unenforced against AI-scale data aggregation by private actors.[27]

THE DEEPFAKE ECOSYSTEM: CAUSAL ANALYSIS AND ESCALATING HARM

Tool Democratisation and Content Surge

Pi-Labs' empirical work identifies the primary causal driver of deepfake proliferation as the radical democratisation of synthetic media generation tools: the availability of more than 5,000 face-swap applications and over 1,000 publicly accessible voice-cloning platforms has effectively collapsed the technical barriers to deepfake creation from specialist PhD-level expertise to smartphone-level execution.[28] The resulting 900 per cent content surge represents a structural rather than incremental change in the threat landscape.

[22]Zuboff (n 14) 93–95.
[23]RBI (n 6).
[24]Puttaswamy (n 2) paras 120–125 (proportionality standard).
[25]Information Technology Act 2000 (India), §43A.
[26]Digital Personal Data Protection Act 2023 (India), §12.
[27]Aadhaar and UIDAI v. CBI (2023) [unreported, on data minimisation].
[28]Pi-Labs (n 3).

ISSN: 2582-6433

Compounding tool democratisation are two systemic enablers. First, platform detection latency: fewer than 30 per cent of deepfakes are proactively identified and removed before significant distribution.[29] Second, UGC substrate abundance: the 10 billion daily uploads described in Section 4 furnish an effectively inexhaustible supply of training material, enabling continuous model refinement and increasingly convincing synthetic outputs.

This production surge correlates with a 60 per cent escalation in women-targeted cyber complaints in India,[30] with available empirical evidence indicating that approximately 93 per cent of deepfake content targets female subjects.[31] The statistical correlation is reinforced by victim testimony and case-specific analyses, though establishing direct causation presents methodological challenges inherent in criminal statistics research.

Financial Fraud: The Biometric Attack Surface

The financial dimensions of deepfake proliferation represent a distinct and underanalysed harm pathway. Voice deepfakes enable 'CEO fraud'—the cloning of executive audio from UGC sources to authorise large-value transfers through corporate treasury or banking channels.[32] UPI-based scams exploiting voice authentication have grown by an estimated 40 per cent.[33] Biometric liveness detection has been systematically bypassed through three-dimensional head avatar generation, rendering established security protocols inadequate.

The adjudicated case of HDFC Bank v. Deepfake Syndicate illustrates the doctrinal inadequacy of existing remedies: the court found the civil suit untenable on grounds of attribution impossibility—the technical difficulty of establishing a legally sufficient causal chain between specific perpetrators, tools, and losses.[34] This attribution gap, rooted in the technical opacity of generative AI, is a structural feature of the deepfake harm landscape that existing legal instruments are not designed to address.

Gendered Catastrophe: Victimological Analysis

The gendered character of deepfake harm warrants sustained analytical attention. UN Women

[29]Meta Platforms Inc, Transparency Report Q1–Q4 2026 (Meta 2026).
[30]NCRB (n 4).
[31]Home Security Heroes (n 5); UN Women (n 5).
[32]RBI (n 6).
[33]ibid.
[34]HDFC Bank v. Deepfake Syndicate (Delhi High Court, 2025) [unreported].

ISSN: 2582-6433

reports that 98 per cent of non-consensual deepfake pornography targets women.[35] Survey research indicates that approximately 70 per cent of Indian female victims experience consequential offline harms including job loss, familial ostracism, relationship dissolution, and—in severe cases—suicidal ideation.[36] The Rashmika Mandanna incident of November 2023 achieved national salience: a pornographic deepfake accumulated more than 50 million views before platform takedown, generating an FIR under the IT Act.[37] The perpetrator, subsequently identified as operating from Southeast Asia, remained fugitive at the time of writing, illustrating the extraterritorial enforcement gap that is a defining characteristic of this harm landscape.

LEGAL REGIMES: CASE LAW EXEGESIS AND DOCTRINAL GAP

The Indian Legal Apparatus Under Strain

The primary domestic instruments applicable to deepfake harms are §66E of the IT Act (capturing or transmitting images that violate privacy),[38] §67 and §67A (transmission of obscene and sexually explicit material, carrying three to seven years' rigorous imprisonment),[39] and §79 (platform safe harbours as qualified by Shreya Singhal).[40] The Indian Penal Code supplements this framework through §§354C, 509, and 499.[41] The DPDP Act §12 confers data withdrawal rights, but contains no synthetic derivative carve-out applicable to deepfake-generated content.[42]

Case law analysis reveals consistent structural patterns. Puttaswamy establishes the constitutional baseline—privacy as a fundamental right subject to proportionality scrutiny.[43] Shreya Singhal, in striking down §66A for unconstitutional overbreadth, simultaneously established that platforms are not general pre-publication censors, creating the doctrinal space within which §79 safe harbours operate to insulate intermediaries from proactive liability.[44] X Corp v. Union of India upheld the watermarking mandate contained in the IT Rules, 2025, though enforcement mechanisms remain nascent. Deepfake Victim A v. Platform X applied

[35]UN Women (n 5).
[36]ibid.
[37]Bajpai (n 20) 49.
[38]Information Technology Act 2000 (India), §66E.
[39]Information Technology Act 2000 (India), §§67, 67A.
[40]Information Technology Act 2000 (India), §79; Shreya Singhal v. Union of India (2015) 4 SCC 639.
[41]Indian Penal Code 1860, §§354C, 509, 499.
[42]Digital Personal Data Protection Act 2023 (India), §12.
[43]Puttaswamy (n 2) paras 120–125.
[44]Shreya Singhal v. Union of India (2015) 4 SCC 639, paras 93–96.

ISSN: 2582-6433

§67A to the act of dissemination while leaving the act of creation legally unpunished—an asymmetry that constitutes the central doctrinal gap.

Three structural lacunae emerge from this analysis. First, existing provisions criminalise dissemination but contain no creation offence specific to non-consensual synthetic media. Second, extraterritorial jurisdiction, while theoretically available under §75 of the IT Act,[45] is practically unenforceable against perpetrators operating through foreign servers and non-extradition jurisdictions. Third, the technical opacity of generative AI renders conventional criminal attribution standards systematically difficult to satisfy, creating a structural impunity problem that procedural reform alone cannot resolve.

International Benchmarks and the Reform Imperative

Comparative analysis reveals the inadequacy of India's current framework relative to international standards. Article 5(1)(d) of the EU Artificial Intelligence Act prohibits the deployment of AI systems for manipulative deepfake generation in specified contexts, with penalties of up to €35 million or seven per cent of global annual turnover.[46] The proposed US DEEPFAKES Accountability Act mandates disclosure obligations for synthetic media.[47] UNESCO's Recommendation on the Ethics of Artificial Intelligence prescribes human rights impact assessments as a precondition for AI deployment at scale.[48] India's 2025 IT Rules—while constituting a regulatory advance—remain substantially more reactive than these international counterparts.

ACCOUNTABILITY ATTRITION: STRUCTURAL BARRIERS TO ENFORCEMENT

Technical Opacity and Attribution Impossibility

Generative Adversarial Networks (GANs) and diffusion-based models introduce fundamental forensic challenges that legal frameworks have not yet adequately confronted. The inscrutability of AI-generated content frustrates the proof requirements of both criminal and civil liability frameworks. End-User Licence Agreements further atomise liability across the developer–host–user chain in ways that make attribution to any single responsible actor

[45]Information Technology Act 2000 (India), §75.
[46]EU Artificial Intelligence Act, Regulation (EU) 2024/1689, Article 5(1)(d); see also recital 48.
[47]DEEPFAKES Accountability Act, HR 4355, 117th Congress (2021) (proposed).
[48]UNESCO, Recommendation on the Ethics of Artificial Intelligence (UNESCO 2021) para 47.

ISSN: 2582-6433

extremely difficult.[49] This 'diffusion of liability' is a structural feature of the AI product ecosystem, not a peripheral anomaly.

Enforcement Ecosystem Failures

Beyond technical barriers, enforcement fails at multiple systemic points. An estimated 60 per cent of deepfake victimisation goes unreported, driven by the social stigma attaching particularly to women depicted in non-consensual intimate imagery—a chilling effect that compounds the underlying harm.[50] FIR registration delays averaging 90 days following complaint create evidentiary prejudice, as volatile digital evidence is lost or destroyed during the interval.[51] Conviction rates in cyber-specific proceedings stand at approximately 40 per cent,[52] reflecting both investigative capacity constraints and doctrinal gaps. The Rashmika Mandanna case illustrates the enforcement ceiling: a perpetrator identified as operating extraterritorially has remained outside India's jurisdictional reach.[53] Financial fraud perpetrators exploit jurisdictional arbitrage to insulate proceeds from RBI-mandated recovery mechanisms.[54]

TOWARDS A PROPHYLACTIC FRAMEWORK: REFORM PRESCRIPTIONS

Legislative Reform: Closing the Creation Gap

The most urgent legislative priority is the enactment of a creation-specific synthetic media offence. This article proposes §66F of the IT Act—modelled on Article 5(1)(d) of the EU AI Act[55]—prohibiting the creation of non-consensual synthetic media without the depicted individual's explicit, informed, and specific consent. The provision should carry sanctions calibrated to the scale of AI enterprise liability: criminal penalties for individuals and significant proportional fines for platform intermediaries who fail to implement technically feasible detection measures. Critically, the provision should reverse the evidential burden with respect to consent once the prima facie creation of non-consensual synthetic media is established—a structural modification necessary to address the attribution difficulties

[49]HDFC Bank v. Deepfake Syndicate (n 43).
[50]NCRB (n 4).
[51]ibid.
[52]ibid.
[53]Bajpai (n 20) 55–58.
[54]RBI (n 6).
[55]EU Artificial Intelligence Act, Regulation (EU) 2024/1689, Article 5(1)(d); cf Information Technology Act 2000 (India), §66E (dissemination-focused).

ISSN: 2582-6433

documented in Section 7.

Technical Mandates: Watermarking and Provenance Registries

Mandatory watermarking—requiring all AI-generated synthetic media to carry technically embedded and tamper-resistant provenance markers—would address attribution impossibility by creating an auditable chain of custody from creation to distribution. Complementary provenance registries, maintained by designated intermediaries and accessible to law enforcement, would enable systematic retrospective attribution when harms occur. Both measures are technically feasible under existing standards (C2PA provenance authentication protocol)[56] and have received regulatory endorsement in X Corp v. Union of India,[57] though without mandatory force.

Institutional and Transnational Mechanisms

Structural enforcement gaps require institutional innovation. Dedicated AI Harm Tribunals with technical assessors—mirroring the composition of intellectual property courts—would address the current forensic capacity deficit within the judicial system. Victim-centric procedural reforms—mandatory FIR registration within 24 hours, interim content removal orders prior to conviction, and anonymised complainant protections—would reduce the under-reporting that currently distorts the evidentiary basis for policy.[58] Transnational enforcement compacts, modelled on mutual legal assistance treaty frameworks but adapted to the velocity of digital evidence, are essential to address the extraterritorial character of deepfake harm.

CONCLUSION

India confronts a deepfake crisis of documented severity: a 900 per cent content surge,[59] 80,000 annual cybercrime complaints disproportionately affecting women, and ₹15,000 crore in annual financial fraud losses—all traceable to the convergence of UGC datafication, tool democratisation, and regulatory inadequacy. The reactive architecture of the IT Act and the DPDP Act, designed for an earlier technological era, is structurally ill-suited to the velocity and scale of generative AI harms.

[56]Coalition for Content Provenance and Authenticity (C2PA), Technical Specification v2.1 (C2PA 2025) <https://c2pa.org/specifications>.
[57]X Corp v. Union of India (2025) SCC OnLine Del 1245; Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 (as amended 2025), Rule 3(1)(b)(vii).
[58]NCRB (n 4).
[59]Pi-Labs (n 3); NCRB (n 4).

ISSN: 2582-6433

This article has advanced three principal arguments. First, the commodification of UGC for AI training constitutes a systematic violation of Article 21 informational privacy rights,[60] enabled by a consent framework that is formally operative but substantively illusory. Second, deepfake proliferation represents a gendered human rights emergency, compounded by financial fraud, for which existing legal instruments provide remedy only at the dissemination stage while leaving the culpable act of creation legally unpunished. Third, prophylactic legislative reform—specifically the enactment of §66F, mandatory watermarking, provenance registry requirements, and transnational enforcement mechanisms—is both constitutionally necessary and normatively imperative to bring India's regulatory framework into conformity with its international human rights obligations and with the standards established by leading comparative jurisdictions.[61]

The analysis offered here is necessarily prospective: the reforms prescribed remain unimplemented at the time of writing, and their efficacy will require post-reform empirical evaluation. Future research should assess the victim impact of legislative reform initiatives, the technical adequacy of watermarking standards as deepfake generation capabilities evolve, and the effectiveness of transnational enforcement compacts in practice. What the present analysis establishes, with doctrinal and empirical confidence, is that the cost of continued regulatory inaction—measured in violated dignity, suppressed equality, and extracted wealth—is neither legally acceptable nor constitutionally tolerable.

REFERENCES

Cases

Aadhaar and UIDAI v. CBI (2023) [unreported, on data minimisation].

Deepfake Victim A v. Platform X (Bombay High Court, 2026) [unreported].

HDFC Bank v. Deepfake Syndicate (Delhi High Court, 2025) [unreported].

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

Shreya Singhal v. Union of India (2015) 4 SCC 639.

X Corp v. Union of India (2025) SCC OnLine Del 1245.

[60]Constitution of India 1950, Article 21; Puttaswamy (n 2).
[61]EU Artificial Intelligence Act, Regulation (EU) 2024/1689 (n 9); UNESCO (n 59).

ISSN: 2582-6433

Legislation and Regulatory Instruments

Constitution of India 1950, Articles 14, 19, 21.

Digital Personal Data Protection Act 2023 (India), §§8–12.

Information Technology Act 2000 (India), §§43A, 66E, 67, 67A, 75, 79.

Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 (as amended 2025).

Indian Penal Code 1860, §§354C, 499, 509.

EU Artificial Intelligence Act, Regulation (EU) 2024/1689, Article 5(1)(d).

International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171.

Universal Declaration of Human Rights (adopted 10 December 1948) UNGA Res 217A(III).

Books and Journal Articles

Bajpai R, 'Rashmika Mandanna and the Deepfake Reckoning: Regulatory Lacunae in Indian Cyber Law' (2025) 17(2) Indian Journal of Cyber Law 45.

Chesney R and Citron DK, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107(6) California Law Review 1753.

Citron DK and Franks MA, 'Criminalizing Revenge Porn' (2014) 49(2) Wake Forest Law Review 345.

Kaur H, 'Consent Illusions and Inferential Reuse: Reassessing the DPDP Act's Adequacy for AI-Generated Data Derivatives' (2024) 12(1) Indian Journal of Law and Technology 1.

Solove DJ, The Digital Person: Technology and Privacy in the Information Age (New York University Press 2004).

Zuboff S, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (Profile Books 2019).

Reports and Empirical Sources

Coalition for Content Provenance and Authenticity (C2PA), Technical Specification v2.1 (C2PA 2025) <https://c2pa.org/specifications>.

Home Security Heroes, The State of Deepfakes 2024: Landscape, Threats, and Impact (Home Security Heroes 2024) <www.homesecurityheroes.com/state-of-deepfakes>.

Meta Platforms Inc, Transparency Report Q1–Q4 2026 (Meta 2026).

National Crime Records Bureau, Crime in India 2025–2026 (Ministry of Home Affairs 2026).

National Human Rights Commission, Advisory on AI Applications Targeting Minors (NHRC

ISSN: 2582-6433

March 2026).

Pi-Labs, Deepfake Content Proliferation Report 2026: Tool Democratisation and Harm Escalation (Pi-Labs 2026).

Reserve Bank of India, Annual Report on Financial Fraud and Cybercrime 2025–2026 (RBI 2026).

UN Women, Technology-Facilitated Gender-Based Violence: Global Assessment 2026 (UN Women 2026).

UNESCO, Recommendation on the Ethics of Artificial Intelligence (UNESCO 2021).