Jus Corpus Law Journal

Open Access Law Journal – Copyright © 2026 – ISSN 2582-7820

Editor-in-Chief – Prof. (Dr.) Rhishikesh Dave; Publisher – Ayush Pandey

The Legal Vaccum around Deepfake Abuse: Gender, AI, and Liability in India

Ramya Prakash[a] NS Nidhi[b]

[a]Assistant Professor, School of Law, M S Ramaiah University of Applied Sciences, Bangalore, India

[b]School of Law, M S Ramaiah University of Applied Sciences, Bangalore, India

Received 17 January 2026; *Accepted* 20 February 2026; *Published* 27 February 2026

The rapid growth of generative AI has made it increasingly easy to produce non-consensual intimate deepfakes, intensifying gender-based online harassment while also fueling broader misinformation, disinformation, and malinformation (MDM) campaigns. These synthetic images circulate rapidly across digital platforms, yet India's legal framework remains fragmented. Provisions of Bharatiya Nyaya Sanhita (BNS), Information Technology (IT) Act, Protection of Children from Sexual Offences Act (POCSO), and Prevention of Sexual Harassment (POSH)Act are not very clear in terms of assigning liability for AI-generated content and providing effective redressal for victims. The draft IT Intermediary Rules 2025 on Synthetically Generated Information (SGI), which aim at compulsory labelling, permanent metadata, user declarations, and verification mechanisms, show a certain level of advancement but still do not quite address the gendered harms aspect of the intimate deepfakes. The Digital Personal Data Protection Act (DPDPA) Rules 2025 lay down the definition of Significant Data Fiduciaries (SDF) as those entities that handle personal data on a large scale. Consequently, they are obligated to put in place protective measures that would prevent the misuse of algorithms. This is a direct way of dealing with the problem of the liability gap in tort and criminal law by stating that AI companies should not allow their systems to be used as weapons to create non-consensual intimate deepfakes of women. This study examines doctrinal gaps in tort, criminal, and evidentiary law and advocates for reforms based on constitutional guarantees of equality and dignity. Such a cohesive, AI-aware regulatory framework would be necessary to secure accountability, increase the level of protection, and address in a substantive way the harassment that is enabled by deepfakes as well as the manipulation of information.

Keywords: *deepfake abuse, gendered online harassment, AI liability, synthetic media regulation, India's legal framework.*

INTRODUCTION

The worldwide spread of generative AI in a very short span has changed the face of digital threats considerably. Women and marginalised communities in India are most affected by such threats. Deepfake is an AI-powered synthetic media that can manipulate faces, voices, and identities with a scary level of similarity. It has become a major tool for gender-based violence, harassment, and disinformation.[1] Non-consensual intimate deepfakes that put someone's face on a sexually explicit video without their consent are, perhaps, the most evil side of this technology, causing the victims, mostly women, immense psychological, reputational, and economic harm.[2] This paper identifies three interrelated legal vacuums: (a) a doctrinal gap: no statute explicitly criminalizes non-consensual intimate deepfakes as a distinct harm; (b) a liability gap: unclear responsibility across creators, distributors, platforms, and AI-tool providers; and (c) an evidentiary and procedural gap: no standards for authenticating synthetic media or providing rapid, victim-centred remedies.

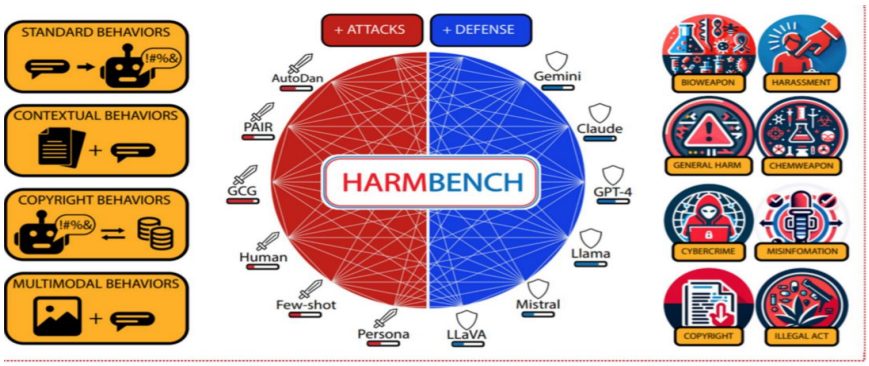
India's constitution lays down the rules that also necessitate consideration of fast changes. Article 14 guarantees equal treatment under the law and thus calls for special steps to eradicate gender-based discrimination. Article 15 forbids discrimination based on sex. Article 21, as held in the K S Puttaswamy v Union of India[3], the right to privacy is among the fundamental rights. Privacy rights thus entail being the sole decision-maker on one's use of one's image, likeness, and informational autonomy. Articles 19(1)(a) and 19(2) deal with free speech, and such speech is subject to reasonable restrictions in the interests of decency, morality, and preservation of an individual's rights. Deepfake measures, if they are just about being narrowly targeted at gendered harms and not used for political censorship, would be covered by these allowed restrictions, hence making reform not a simple policy choice but a constitutional necessity.

---

[1] 'Legal Dimensions of Deepfake Technology: Privacy, Consent and Criminal Liability' (*Juris Centre,* 27 July 2025) <https://juriscentre.com/2025/07/27/legal-dimensions-of-deepfake-technology-privacy-consent-and-criminal-liability/> accessed 19 November 2025

[2] 'Deepfake trends and challenges - position statement' (*eSafety Commissioner*) <https://www.esafety.gov.au/industry/tech-trends-and-challenges/deepfakes> accessed 20 November 2025

[3] *Justice K S Puttaswamy (Retd) & Anr v Union of India & Ors* (2017) 10 SCC 1



9 Workflow of the HarmBench framework for Automated Red teaming and mitigating misinformation risks (Mazeika et al. 2024)

The present research article delves into the puzzle of legal challenges deepfake misuse poses in India, exploring the aspects of gender discrimination in particular. It identifies significant gaps in tort, criminal, and evidentiary law that are barriers to achieving justice. This is done through doctrinal analysis of the existing legislative provisions, judicial precedents, and regulatory frameworks at the nascent stage. This study moves from these gaps to the provisions in the Constitution related to gender equality under Article 14 and human dignity under Article 21, as well as to worldwide regulatory approaches.[4] It uses these as the basis for the paper's argument for the adoption of a comprehensive legal reform framework, mindful of AI technology and placing the protection of victims at the centre while considering innovation, freedom of expression and platform accountability.

RESEARCH QUESTIONS

The research combines doctrinal analysis of existing legislative provisions and judicial precedents, comparative analysis of deepfake governance in selected foreign jurisdictions, and policy analysis of India's emerging regulatory frameworks, such as the draft IT Intermediary Rules 2025 and DPDP Rules 2025.

[4] Michael A Santoro and Anonymous, 'Deep Fakes and Surveillance Technology: Comparing the EU AI Act and Chinese AI Regulation' (*Business Human Rights Journal*, 05 February 2025) <https://bhrj.blog/2025/02/05/deep-fakes-and-surveillance-technology-comparing-the-eu-ai-act-and-chinese-ai-regulation/> accessed 15 November 2025

Research Question	Focus Area	Key Legal Domains Involved
RQ1: How sufficient are the current Indian legal frameworks in tort, criminal, and evidentiary law to deal with non-consensual intimate deepfakes and to provide effective remedies to women victims of gendered online harassment?	Adequacy of existing law	Tort law, criminal law (BNS, IT Act, POCSO), evidentiary law i.e Bharathiya Sakshya Adhinayam (BSA), gender-justice frameworks
RQ2: What changes in doctrine and regulations, based on India's constitutionally guaranteed rights of equality and dignity, are required to establish an AI-aware liability regime for deepfake misuse, including more transparent evidentiary standards for synthetic media?	Legal reform needs	Constitutional law (Arts. 14, 15, 19, 21), digital evidence standards, and regulatory reform
RQ3: How can India craft a balanced regulatory framework that simultaneously integrates platform accountability, AI, tool governance, and victim-centric protections to prevent, detect, and remediate non-consensual intimate deepfakes?	Regulatory design	IT Rules 2025, DPDPA 2025, platform liability, algorithmic accountability, victim, rights mechanisms

Table 1: Research Questions

RESEARCH METHODOLOGY

The research combines doctrinal analysis, comparative review, and desk research to understand legal and policy systems' response to intimate deepfakes and the broader

ecosystem. The research takes the first step of systematically reviewing the international academic literature on topics such as AI governance, online safety, data protection, and platform regulation to understand the major debates and determine the gaps, notably those related to gendered harms and accountability.

Alongside, the investigation relies on the vast variety of secondary sources, such as statutes, policy papers, government publications, international standards, and AI-governance frameworks, most of which are not available in the standard academic databases. This desk-based approach is necessary for understanding a quickly changing regulatory environment, where a large part of the main advice is very often from non-academic entities. The combination of the data from both academic and policy sources makes the research more reliable, globally relevant, and it reflects the actual regulatory debates more accurately.

THE DEEPFAKE CRISIS: UNDERSTANDING THE THREAT

The Technology and Its Evolution: Deepfakes are AI-powered synthetic media that manipulate faces, voices, and identities with uncanny realism that exemplify the misinformation-disinformation-malinformation (MDM) spectrum. Misinformation involves unintentional false information being circulated, while disinformation involves deliberate creation and dissemination, and malinformation is genuine information weaponised maliciously. These have become major tools for gender-based violence, harassment, and targeted silencing of women in digital spaces. These AI models learn from a large number of images, videos, and audio recordings, which allows them to create extremely realistic falsifications that are becoming more and more indistinguishable from genuine ones.[5] The spread of this technology through the applications that are easy to use and the platforms that are available online has removed the obstacles that were present in the field, so now practically anyone who has access to the internet can make a deepfake.[6]

The rate at which the deepfake crisis has been escalating exponentially is still ongoing. Global detection techniques for these harmful files have identified over 8 million deepfake videos in

[5] Dr Ajit Singh and Dr Pawan Kumar, 'Digital Evidence and Deepfake: A Challenge to Criminal Justice System in India' (2025) 12(8) Journal of Emerging Technologies and Innovative Research <https://www.jetir.org/papers/JETIR2508273.pdf> accessed 15 November 2025

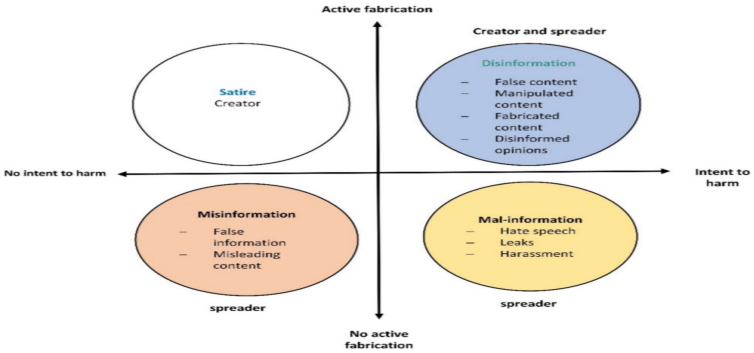
[6] Noelle Martin, 'Online safety regulation of deepfake abuse: a case study of Australia's eSafety Commissioner' (2025) 34(1) Griffith Law Review <https://www.tandfonline.com/doi/full/10.1080/10383441.2025.2504791> accessed 14 November 2025

2024, which is a sixteen times higher number than that of 2020.[7] India has experienced a considerable increase in deepfake occurrences, mainly the ones which are done to women celebrities, activists, and common people.[8] In India, BNS Section 294[9] and IT Act Sections 67/67A[10] apply uniformly across this spectrum without distinguishing harm intent or gendered targeting, creating regulatory blind spots. The cases involving the actresses Rashmika Mandanna and Katrina Kaif, in which their images were misused, brought the problem to the forefront of the nation, thus resulting in the issuing of advisory notices by the government and the taking of preventive measures by the law enforcement agencies.[11]

According to the research, internet deepfakes amounting to 96% are of a pornographic nature without the consent of the involved parties, and the victims of which are women for the most part.[12] A 2024 McAfee survey found 75% of Indians encountered deepfakes, and 38% were personally targeted by deepfake scams, with this concern formally raised in the Lok Sabha in March 2025. The gendered problem referred to at the beginning corresponds to the bigger problem of technology, facilitated abuse and women's online harassment that women and girls, especially those in public life who are against the patriarchy or engage in social justice activities, are mostly targeted by.[13]

Dimensions of Harm: Abuse of deepfake technology, among other things, has deeply violated people's privacy, dignity, and freedom. Without someone's consent, using their image, and most probably, in a pornographic manner, is a grave offence which is aimed at a person's physical and psychological integrity. Victims report that they are powerless, violated, ashamed, and that they lose control over their digital footprint, and this can result

[7] 'AI Generated Content Regulation in India' (*Drishti IAS,* 25 October 2025) <https://www.drishtiias.com/daily-updates/daily-news-editorials/ai-generated-content-regulation-in-india> accessed 21 November 2025
[8] Yash Bajpai, 'Me, Myself and AI: Chasing Deepfakes Across Borders Without Losing Your Rights' *SCC Online* (08 November 2025) <https://www.scconline.com/blog/post/2025/11/08/deepfake-regulation-rights/> accessed 19 November 2025
[9] Bharatiya Nyaya Sanhita 2023, s 294
[10] Information Technology Act 2000, ss 67 and 67A
[11] Shuchi Nagpal, 'Deepfakes & Cyber Law' (*Asian Laws,* 06 February 2024) <https://www.asianlaws.org/blog-post.php?url=deepfakes-and-cyber-law> accessed 11 November 2025.
[12] 'AI, Consent, and Workplace Harassment: Navigating Deepfakes and Digital Misconduct' (*Safe Spaces*) <https://safespacesinc.in/ai-consent-workplace-harassment/> accessed 12 November 2025
[13] Priyanka Kumari et al., 'Digital Privacy At Risk: "Examining India's Legal Response to the Non-Consensual Sharing of Intimate Media"' (2025) 3(3) IJLSSS <https://ijlsss.com/digital-privacy-at-risk-examining-indias-legal-response-to-the-non-consensual-sharing-of-intimate-media/> accessed 20 November 2025



in psychological effects like anxiety, depression, social isolation, and, in extreme cases, suicidal ideation.[14]

Socially, the disadvantages caused by deepfake technology are not limited to individual victims but also affect the communities and democratic institutions. Deepfakes are used in identity weaponisation for harassment, intimidation, and silencing, most of the time targeting women activists, journalists, and public figures who oppose the existing power structures.[15] Political deepfakes threaten the stability of elections by disinforming, faking the identity of the candidates, and hence influencing voters. A few deepfake videos of false political statements and endorsements by influential figures were circulated widely during the 2024 Indian general elections.[16] Economically, the damage occurs when someone fraudulently impersonates, damaging their reputation. The damage to the public image is the primary factor for the change of the women who were the victims of non-consensual intimate deepfakes, to the acknowledgement of career losses, lost employment opportunities, and economic disenfranchisement, is most evident.[17]

[14] Nicola Henry and Gemma Beard, 'Image-Based Sexual Abuse Perpetration: A Scoping Review' (2024) 25(5) Trauma, Violence, & Abuse <https://journals.sagepub.com/doi/10.1177/15248380241266137> accessed 13 November 2025

[15] Gopal Trivedi, 'Deepfakes, Identity, Persona and the Indian Legal Frontier' (*Chand & Chand Intellectual Property*, 18 November 2025) <https://www.candcip.com/single-post/deepfakes-identity-persona-and-the-indian-legal-frontier> accessed 21 November 2025

[16] Vamsi Krishna Pothuru, 'Deepfakes, cloned voices, and digital media literacy: AI's role in the misinformation crisis in India' (*WACC Global*, 13 August 2025) <https://waccglobal.org/deepfakes-cloned-voices-and-digital-media-literacy-ais-role-in-the-misinformation-crisis-in-india/> accessed 19 November 2025.

[17] Dhanya Lakshmi et al., *Digital violence, real world harm: evaluating survivor-centric tools for intimate image abuse in the age of generative AI* (Human Intelligence 2025)

INDIA'S FRAGMENTED LEGAL FRAMEWORK

Constitutional Foundations: India's constitutional system offers initial safeguards that are relevant to deepfake usage through its fundamental rights embodied in Articles 14, 15, 19, and 21. Article 14[18] ensures equality before the law and equal protection of the laws to all people in the Indian territory. This principle forbids the State from making really discriminatory decisions and therefore requires that the State treat persons in the same position equally.

Article 21, which guarantees the right to life and personal liberty, has also been broadly construed by the Supreme Court to include the rights to privacy, dignity, informational autonomy, and bodily integrity.[19] In the landmark judgment K S Puttaswamy v Union of India,[20] privacy was declared a fundamental right, thereby acknowledging that human beings have control over their personal information, image, and likeness. Such a constitutional validation is the normative basis for the legal protection of individuals against the use of their identity without consent in deepfakes.[21]

Article 19(1)(a) is concerned with the protection of freedom of speech and expression, while Article 19(2) allows for restrictions that are reasonable and necessary in the interests of sovereignty, security, public order, decency, morality, defamation, and contempt of court. The constitutional equilibrium between free speech and allowable restrictions becomes an issue of the highest importance in the case of deepfake regulation, where the steps taken to fight disinformation and intimate image abuse should be in the right proportion so as not to result in excessive censorship.[22]

Information Technology Act 2000: The Information Technology (IT) Act 2000 is the main legal framework of India, which regulates cybercrimes and electronic communications.

---

[18] Indian Constitution 1950, art 14

[19] Krunal Mehta, 'The Persona Paradox: Personality Rights, Deepfakes & Identity in Indian Law' (*King Stubb & Kasiva,* 06 October 2025) <https://ksandk.com/media-and-entertainment/the-persona-paradox-deepfakes-personality-rights-in-india/> accessed 20 November 2025

[20] *Justice K S Puttaswamy (Retd) & Anr v Union of India & Ors* (2017) 10 SCC 1

[21] Khushi Saraf and Akshay Sriram, 'The Dilemma of Deepfakes: Expanding the Ambit of Right to Personality to Regulate Deepfakes in India' (*Law School Policy Review*) <https://lawschoolpolicyreview.com/2024/05/04/the-dilemma-of-deepfakes-expanding-the-ambit-of-right-to-personality-to-regulate-deepfakes-in-india/> accessed 18 November 2025

[22] Divij Joshi, 'Intermediary Rules 2021: The Constitution and Content Moderation (Part II)' (*Centre for Law and Policy Research,* 03 April 2021) <https://clpr.org.in/blog/intermediary-rules-2021-the-constitution-and-content-moderation-part-ii/> accessed 11 November 2025

Several of its provisions extend to offences related to deepfakes. Section 66C enacts a penalty for identity theft through electronic means.[23] Section 66D makes cheating by personation in computer resources an offence punishable with a jail term of up to three years and a fine of up to one lakh rupees.[24] Both these provisions focus on cheating or deception, but may not cover all deepfake offences. This is because many intimate deepfakes are created to humiliate or intimidate women rather than to obtain money or property, so they fall outside the economic-fraud logic underlying Section 66D and Section 336 of BNS[25]. As a result, the law treats all obscene or sexually explicit electronic material in broadly similar terms and does not conceptually distinguish consensual synthetic imagery from non-consensual intimate deepfakes, even though the lack of consent is the central harm in such cases.

Section 66E deals with privacy issues and penalises the unauthorised capturing, publication, or transmission of private pictures with imprisonment of up to three years or a fine of up to two lakh rupees.[26] Sections 67, 67A, and 67B are mainly about the prohibition of the publication and transmission of electronically formatted obscene, sexually explicit, or child sexual abuse materials. In a case, the petitioner invoked 67 and 67A against the circulation of morphed images as an offence to public decency and morality. There was limited coverage of the lack of consent in deepfake offences. In relation to the just stated, DPDPA is yet to implement the right to be forgotten as a data right, like in the General Data Protection Regulation (GDPR), a crucial tool in protecting reputational damage online.[27] Section 67B talks about the Child Sexual Abuse material (CSAM), including computer-generated images depicting children in sexually explicit acts. The 2024 Supreme Court ruling very emphatically states that, except in very narrowly defined situations, possession and viewing of CSAM are crimes. Thus, the number of people who can be held liable has increased considerably.[28]

[23] Information Technology Act 2000, s 66C
[24] Information Technology Act 2000, s 66D
[25] *Ibid*; Bharatiya Nyaya Sanhita 2023, s 336
[26] Information Technology Act 2000, s 66E
[27] Information Technology Act 2000, ss 67, 67A and 67B; *Google Spain SL v Agencia Espanola de Proteccion de datos* [2014] C-131/12
[28] *Just Rights for Children Alliance & Anr v S Harish & Ors* (2024) INSC 716; Prachi Bhardwaj, 'Storing and watching child pornography an offence under POCSO, IT Act: Key takeaways from Supreme Court's Landmark Verdict' *SCC Online* (23 September 2024) <https://www.scconline.com/blog/post/2024/09/23/storing-watching-child-pornography-crime-supreme-court-pocso-it-act/> accessed 15 December 2025

Section 79 provides conditional safe harbour immunity to intermediaries for third-party content. It means that intermediaries will not be held liable if they perform their due diligence and, after obtaining actual knowledge, remove the illegal content without delay.[29] The Supreme Court's judgment in Shreya Singhal v Union of India[30] made it clear that the essential intermediaries' protections mean that they can only be requested to remove content upon court orders or government notifications on legitimate grounds under Article 19(2).

The Delhi High Court has taken judicial notice of deepfake threats in Rajat Sharma v Union of India and Chaitanya Rohilla v Union of India, pressing the government to constitute a committee to examine the dangers posed by deepfakes to privacy and public trust; however, no statutory response has materialised.[31]

Bharatiya Nyaya Sanhita 2023: Many of the provisions apply to deepfake crimes in the Bharatiya Nyaya Sanhita (BNS) that replaced the Indian Penal Code in 2024. As a case in point, Section 336 is about cheating through personation, and Section 356 is a forgery-related crime that also involves digital forgery of pictures and documents.[32] Section 294 is concerned with preventing the publication and transmission of any electronic obscene material, and this applies to deep fake pornography, as well.[33]

Nevertheless, deepfakes are not defined by the BNS as a distinct type of crime, and, therefore, there exist certain loopholes to the potential damages of AI that the legislation does not acknowledge. Consequently, prosecutors must fit non-consensual intimate deepfakes into general offences on cheating, forgery, or obscenity rather than a dedicated image-based abuse offence, which weakens the clarity, consistency, and signalling value of the law. The National Commission for Women has suggested that the regulations should be changed

[29] Information Technology Act 2000, s 79
[30] *Shreya Singhal v Union of India* (2015) 5 SCC 1
[31] 'Pixelated Perjury: Addressing India's Regulatory Gaps in Tackling Deepfakes' (*Tech Law Forum NALSAR*, 18 December 2025) <https://techlawforum.nalsar.ac.in/pixelated-perjury-addressing-indias-regulatory-gaps-in-tackling-deepfakes/> accessed 18 December 2025
[32] Aditi Pangotra, 'Countering Misinformation: Provisions under the Bharatiya Nyaya Sanhita, 2023' (*Cyber Peace,* 09 September 2024) <https://www.cyberpeace.org/resources/blogs/countering-misinformation-provisions-under-the-bharatiya-nyaya-sanhita-2023> accessed 14 November 2025
[33] Bharatiya Nyaya Sanhita 2023, s 294

accordingly to ensure legal definitions of modified content are clear and that there are precise penalties for violating deepfakes in the BNS.[34]

POCSO Act 2012: The Protection of Children from Sexual Offences (POCSO) Act 2012 is a law that defines various harmful acts against children sexually, including the manufacture, distribution, and possession of child sexual abuse material (CSAM).[35] In its 2024 ruling, the Supreme Court has substantially extended the application of POCSO. It has been held that 'mere possession' is a penal provision and that AI-generated CSAM with identifiable children is a matter under the Act.[36] The Court explained that changed or altered images that seem to show children, including deepfakes, are considered to meet the legal provision. This explicit inclusion of AI-generated CSAM reveals a protection asymmetry: synthetic sexual images of children are clearly covered, whereas synthetic intimate images of adults, overwhelmingly targeting women, lack parallel explicit statutory recognition.

POSH Act 2013: The Prevention of Sexual Harassment (POSH) Act 2013 requires the establishment of safeguards in the workplace against any constitutive part of sexual harassment, which has to be extended to virtual and remote work settings as well.[37] Even if the law does not pinpoint the issue of deepfakes specifically, judges and lawyers have interpreted its various provisions as implying that there can be cases of harassment which happen digitally and that these can be crimes that happen by using means such as WhatsApp, e-mail, video calling, and social media. In fact, the Delhi High Court in 2025 ruled that sending offensive messages via Facebook and WhatsApp amounts to sexual harassment at the workplace under the POSH Act; thus, even if the incidents take place in non-office areas, they are still covered by the Act.[38]

Nevertheless, there are still considerable differences in the way POSH can be used to deal with problems arising from AI-powered harassment. The majority of Internal Committees

---

[34] Ambika Pandit, 'NCW recommends legal definition, penalties under criminal law to counter deep fake *abuse*' *The Times of India* (11 November 2025) <https://timesofindia.indiatimes.com/india/ncw-recommends-legal-definition-penalties-under-criminal-law-to-counter-deep-fake-abuse/articleshow/125241763.cms> accessed 20 November 2025

[35] Protection of Children from Sexual Offences Act 2012, ss 14 and 15

[36] *Just Rights for Children Alliance & Anr v S Harish & Ors* (2024) INSC 716

[37] Shagun, 'Sexual Harassment in Virtual Workplaces: How the POSH Act Covers Online Harassment.' (*Indian Working Woman,* 02 April 2025) <https://indianworkingwoman.org/node/19> accessed 11 November 2025

[38] Arjun Paleri, 'Facebook, Whatsapp as Workplace: Court Reaffirms PosH Act's Protection in Virtual Spaces' (*BTG Advaya,* 08 August 2025) <https://www.btgadvaya.com/post/facebook-whatsapp-as-workplaces-court-reaffirms-posh-act-s-protection-in-virtual-spaces> accessed 12 November 2025

(ICs) are not equipped with the necessary knowledge to recognise various situations of deepfake-induced sexual harassment, especially in cases when the abusers are anonymous or when the use of unauthorised channels is involved.

Taken together, constitutional guarantees and sector-specific statutes offer only a fragmented response to deepfake abuse. Criminal law stretches pre-AI concepts like cheating, obscenity and privacy; civil law provides underdeveloped and expensive remedies; and intermediary rules emphasise safe harbour more than rapid, victim-centred takedown. These structural weaknesses collectively constitute the legal vacuum around non-consensual intimate deepfakes in India.

GENDERED DIMENSIONS OF DEEPFAKE ABUSE

Disproportionate Targeting of Women: One after another, studies show that women are the group that suffers most from deepfake abuses. The research result expresses that 96% of deepfake pornographic materials are about women, where female celebrities, female politicians, female activists, and average women are the targets of tech-facilitated gender-based violence. The gendered pattern not only mirrors but also supports patriarchal power structures that try to silence, shame, and control women through sexualized violence.[39] This pattern aligns with the general findings of studies on technology, facilitating gender-based violence. It has been revealed that women who are involved in public life and those who are challenging patriarchal norms are the main targets of online abuse.

Among the examples of the targeting of high-profile women are the case of the actress Rashmika Mandanna and the case of the activist Kamya Buch. In the Rashmika Mandanna incident, a non-consensual intimate deepfake of the actress was circulated widely across platforms, triggering public outrage and a criminal investigation, yet also exposing the absence of any deepfake-specific offence in Indian law. Kamya Buch, a young activist, faced deepfake threats aimed at discrediting her advocacy, illustrating how such abuse is used to punish women for political speech and social justice work. Women are subjected to deepfake attacks as a part of the misogynistic harassment that is being used as a weapon by deepfakes

[39] 'Algorithms and accountability: rewiring artificial intelligence for gender equality' (*EIGE*, 14 April 2025) <https://eige.europa.eu/newsroom/director-corner/algorithms-and-accountability-rewiring-artificial-intelligence-gender-equality> accessed 16 November 2025

to persecute women for their visibility, achievements, and advocacy.[40] Girls going out and telling people about social issues, criticising unfair treatments, or being in positions of influence, in fact, face the most risk of being met by deepfake reactions, which are aimed at discrediting and intimidating them.[41]

The gendered characteristic of deepfake abuse is also intertwined with various other factors leading to discrimination, like caste, religion, class, and sexuality. Several in-depth reports have unveiled the deliberate production of AI pornographies that depict the scenario in which Muslim women have sexual relations with Hindu men. Such implementation of deepfakes is a weapon to spread communal hatred and empower patriarchal dominance. These communalised deepfakes show that gendered deepfake abuse does not operate in isolation, but intersects with religious and caste hierarchies to reproduce existing power structures in more virulent digital forms.

Psychological and Economic Consequences: The psychological consequences of intimate deepfakes that are non-consensual are frighteningly powerful and lasting for a long time. The traumatising experiences recounted by the victims in many cases parallel those of sexual assault, as the victims have been found to develop symptoms of anxiety, depression, post-traumatic stress disorder, withdrawal from social life, and even thinking about suicide. The fact that online content can be permanent and viral makes these damages even more severe. Different from physical assaults that take place locally and temporally, the abuse of intimate images continues forever in the digital world, and thus, victims are consecutively exposed to trauma from different times when the content is being shared or viewed.

The financial implications are extremely harsh for women. It is the professional reputations that are the main victims of explicit deepfakes when they go viral on the internet, and as a consequence, there is loss of employment, stagnation in the career, as well as lessened

[40] 'Deepfakes and AI-Generated Content: A Landmark Ruling by the Delhi High Court in Response to a Harassment Campaign Targeting a Public Activist' (*DDG*, 21 July 2025) <https://en.ddg.fr/actualite/deepfakes-and-ai-generated-content-a-landmark-ruling-by-the-delhi-high-court-in-response-to-a-harassment-campaign-targeting-a-public-activist> accessed 21 November 2025

[41] Sakshi, 'Delhi High Court Grants Interim Relief to Woman in Deepfake Defamation Case, Orders Takedown of Explicit Content' *Law Beat* (18 July 2025) <https://lawbeat.in/top-stories/delhi-high-court-grants-interim-relief-to-woman-in-deepfake-defamation-case-orders-takedown-of-explicit-content-1513414> accessed 22 November 2025

economic possibilities. Women working in client-facing roles, education, media, and public sectors are, therefore, more susceptible to losing their careers as a result of deepfake attacks.[42]

Chilling Effects on Participation: The possibility and actual misuses of deepfakes have frightened and impacted the women so hard that they have created 'chilling effects' in the women's participation in public life, online spaces, and democratic discourse. Women activists, journalists, and public figures, who have experienced such effects, say that they alter their statements, limit their online interactions, and cease using public platforms to avoid being targeted by deepfakes. Recent survey-based research on women journalists and activists in India indicates that a significant proportion self-censor online due to fear of deepfake targeting, reporting reduced posting, withdrawal from contentious debates, and avoidance of visual media such as videos and live streams. The chilling effect here is at odds with the objectives of gender equality and the principles of democracy because it leads to the vanishing of women's voices, which happen to be the most engaged ones in public life issues.[43]

The gendered pattern of deepfake abuse, its disproportionate psychological and economic impacts on women, its intersection with caste and communal hierarchies, and its chilling effect on women's public participation demonstrate that deepfakes are not a neutral cybercrime problem. They function as tools of gender-based violence and social control. Any legal response that reacts to deepfakes simply as a generic form of misinformation or obscenity will thus be insufficient. A thorough deepfake regulation should definitely make women's equality, autonomy, and dignity its major normative commitments.

EMERGING REGULATORY RESPONSES

IT Intermediary Rules 2025: Synthetic Media Regulation: The draft rules for the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules 2025, which were unveiled in October 2025, represent India's greatest

[42] Ankita Deshkar, 'When workplace harassment goes online: PoSH in the remote work era' *The Indian Express* (22 August 2025) <https://indianexpress.com/article/technology/when-workplace-harassment-goes-online-posh-in-the-remote-work-era-10205062/> accessed 11 November 2025

[43] 'Algorithmic Accountability' (*Women at the Table*) <https://www.womenatthetable.net/research-advocacy/algorithmic-accountability/> accessed 10 November 2025

regulatory response to deepfakes to date.[44] These changes define for the first time the term 'synthetically generated information' (SGI) and extend a broad range of obligations, like labelling, identification, verification, and due diligence on the part of the intermediaries and content generators. The Draft SGI Rules represent progress by the introduction of labelling of AI-generated content:

- metadata retention,
- user declarations,
- and verification mechanisms.

However, gaps remain as there is no Content-neutral regulation: Intimate deepfakes are treated the same as political satire or parody; Transparency without accountability: Knowing content is synthetic does not mitigate sexual humiliation; No enhanced duty for amplification: Platforms are not held responsible for how far harmful content spreads.

The proposed regulations require that all artificially generated content bear a prominent, permanent label or an embedded metadata identifier of at least 10% of the visual display area or the audio duration.[45] Significant Social Media Intermediaries (SSMIs), i.e., platforms with more than 5 million users registered in India, are subject to increased liabilities. Among other things, they should ensure that users declare whether the content they upload is synthetically generated and use automated verification tools to check the accuracy of such declarations.

Generator services are those entities that provide computer resources that enable the creation or modification of synthetic media. They are required to attach a permanent identifier at the point of generation. Such an obligation goes beyond AI tools, software, and platforms, by holding technology providers responsible for supplying them on the upstream side.

It is argued that, although these measures go a long way, problems with the implementation process can be discerned by opponents. The accuracy of the detection is still in question because AI detection tools, at present, have both high false positive and false negative rates.

---

[44] 'IT ministry proposal targets synthetic media labeling' (*Ai CERTs*) <https://www.aicerts.ai/news/it-ministry-proposal-targets-synthetic-media-labeling/> accessed 16 November 2025

[45] Shradha Prakash, 'Summary: Draft Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules 2025' (*ALG India,* 17 November 2025) <https://www.algindia.com/summary-draft-information-technology-intermediary-guidelines-and-digital-media-ethics-code-amendment-rules-2025/> accessed 19 November 2025

The requirement of the 10% watermark can make users' experience worse, and it can also be easily circumvented by some technical manipulations. Furthermore, the international distribution of content makes it difficult for enforcement to be consistent since content produced outside of India may not meet the requirements for labelling.[46]

DPDP RULES 2025: ALGORITHMIC ACCOUNTABILITY

The Digital Personal Data Protection (DPDP) Rules 2025 came into effect in November 2025 and lay out in detail data protection obligations that significantly affect the prevention of deepfakes.[47] The Digital Personal Data Protection Act is a significant change in the regulatory framework for Indian platforms as it shifts the focus of responsibility from individual content creators to the entities that design and deploy data- driven systems at a large scale. In the context of intimate deepfakes, the DPDP framework recognises that harm originates not merely in publication but in unlawful data processing, algorithmic amplification, and foreseeable misuse. By imposing heightened obligations on Significant Data Fiduciaries, the Act provides a preventive, dignity-based approach capable of addressing the gendered dimensions of algorithmic harm that criminal and intermediary laws fail to capture. The rules define Significant Data Fiduciaries (SDFs) as those entities that deal with large-scale personal data or data of such sensitivity that it can give rise to significant privacy risks.

Among the SDFs must include various security measures such as encryption, access controls, continuous monitoring, breach notification, and data minimisation.[48] What is more, the DPDP Rules put the SDFs in the position where they are accountable for algorithmic accountability. Firms are mandated to take technical systems, which include algorithmic tools for hosting, displaying, uploading, or sharing personal data, and implement them in a way that they do not become the source of risks to the rights of Data Principals.[49][50]

SDFs have to conduct Data Protection Impact Assessments (DPIAs), analyse fairness, transparency, and objectivity in algorithmic mechanisms and keep records of their

[46] 'Limits of Labelling: Unpacking India's Pursuit to Combat 'Synthetically Generated Information' (*Medianama,* 19 November 2025) <https://www.medianama.com/2025/11/223-limits-labelling-india-combat-synthetically-generated-information/> accessed 20 November 2025
[47] Digital Personal Data Protection (DPDP) Rules 2025
[48] Anahad Narain, 'DPDP Compliance for Large Enterprises' (*Consentin,* 04 July 2025) <https://www.consent.in/blog/significant-data-fiduciary> accessed 11 November 2025
[49] Digital Personal Data Protection (DPDP) Rules 2025
[50] Digital Personal Data Protection Act 2023, s 2(j)

compliance annually.[51] Through these requirements for algorithmic transparency, India ranks among the countries endorsing AI accountability. The DPDP framework does provide for large fines in the event of a violation; the utmost penalties can be as high as ₹250 crore for security failure cases.

Government Advisories and Enforcement: Along with the dependency on a formal regulatory framework, the government of India has also very actively issued multiple advisories to intermediaries concerning deepfake content. Back in November 2023, the Ministry of Electronics and Information Technology (MeitY) instructed the removal of deepfakes from the platform within 36 hours of the notification. To this end, it considered such content as violations of rights and singled out women as those who would be most severely affected.[52]

The police have resorted to criminal charges in the cases of deepfakes that have attracted a lot of public attention. In this regard, as a result, the arrest of Eemani Naveen for fabricating the Rashmika Mandanna deepfake video in November 2023 was a major enforcement turning point.[53] Nonetheless, the execution of the law is still patchy and only happens as a response to the incidents that have already taken place. There are reportedly numerous incidents of deepfakes in which victims are not daring enough to disclose their experience, and the police are not knowledgeable or have limited resources to expose such cases.

Incident Reporting, Victim Rights, and the Limits of Executive Cyber Governance: While the Digital Personal Data Protection Act, 2023, marks an important shift toward recognising individual rights and systemic accountability in data governance, India's cyber incident reporting framework continues to operate in a largely state-centric and compliance-oriented manner. The reporting regime administered by CERT-In prioritises rapid disclosure of cyber incidents to governmental authorities but remains silent on the rights of individuals whose

---

[51] Advocate (Dr.) Prashant Mali, 'DPDP Rules 2025 - Analysis of its implications on Industry and Compliance Guidance' (*DPDPA*) <https://dpdpa.com/blogs/DPDP Rules 2025- Analysis of Industry implications.html> accessed 17 November 2025

[52] 'Government of India Taking Measures To Tackle Deepfakes' (*PIB,* 04 April 2025) <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2119050> accessed 11 November 2025

[53] 'Rashmika Mandanna's deepfake case: Delhi Police close in on suspects; main conspirator remains at large' *Times of India* (20 December 2023) <https://timesofindia.indiatimes.com/entertainment/hindi/bollywood/news/rashmika-mandannas-deepfake-case-delhi-police-close-in-on-suspects-main-conspirator-remains-at-large/articleshow/106147860.cms> accessed 13 November 2025

personal data has been compromised.[54] In contrast, comparative jurisdictions embed breach notification within a rights-based framework: the European Union's General Data Protection Regulation mandates disclosure to affected individuals where a personal data breach is likely to result in risks to their rights and freedoms,[55] while the United States, through its cyber incident reporting framework for critical infrastructure, emphasises timely notification, legislative oversight, and accountability mechanisms.[56] India's failure to impose a uniform statutory requirement on the notification of victims, the publication of breach summaries, or the harmonisation of CERT-In's executive directions with the DPDP Act makes the constitutional right to informational self-determination almost a mere theory.[57] This rift erodes public trust, hides the extent of digital harm, and mainly harms women whose identity violations through data misuse generally lead to irreparable injury to their dignity and reputation. Unless reporting of incidents, victim notification, and procedural safeguards are brought in line with global best practices, India's platform governance regime is likely to give precedence to institutional compliance over the respect of autonomy, equality, and dignity guaranteed by Articles 14 and 21 of the Constitution.[58]

DOCTRINAL GAPS AND EVIDENTIARY CHALLENGES

Doctrinal Gaps: Tort Law: Tort law lacks a codified framework; we rely on common law principles and some statutory remedies. Technically, defamation, invasion of privacy, and intentional infliction of emotional distress may be considered as the causes of deepfakes, but the major doctrinal gaps are hindering the application of civil remedies.[59]

Defamation doctrine is built around propositional falsity: a false statement of fact that harms reputation. Intimate deepfakes complicate this structure because they are visual fabrications rather than explicit statements, yet they still cause severe reputational and dignitary harm. Courts have not yet clarified whether a synthetic video can itself be treated as a defamatory

[54] Information Technology Act 2000, s 70B
[55] General Data Protection Regulation 2016, arts 33 and 34
[56] Cyber Incident Reporting for Critical Infrastructure Act 2022
[57] *Justice K S Puttaswamy (Retd) & Anr v Union of India & Ors* (2017) 10 SCC 1
[58] *Modern Dental College & Research Centre & Ors v State of Madhya Pradesh & Ors* (2016) 7 SCC 353; *Shreya Singhal v Union of India* (2015) 5 SCC 1
[59] Dr Santosh Kumar, 'Legal Implications of Deepfake Technology: Privacy, Consent and Copyright' (2025) 2(1) The Infinite <https://theinfinite.co.in/wp-content/uploads/2025/02/Jan-2025-Legal-Implications-of-Deepfake-Technology.pdf> accessed 15 December 2025

'statement' within the meaning of Section 499 BNS, or whether victims must frame their claims indirectly through accompanying captions or comments.

The doctrines of defamation are challenged by the case of deepfakes, which represent events that did not happen, thus raising the question of whether defamation law should be applied and, if so, in what way damages should be calculated.[60] The requirement of falsity is difficult to meet when deepfakes take some aspects of reality and fabricate others, or show a logical scenario, but the viewers cannot verify it. In addition, the identification and the location of the anonymous deepfake creators make it practically very difficult to file civil claims against them. This anonymity and cross-border dissemination also make civil defamation actions practically ineffective for many women victims, who cannot identify a solvent defendant in India against whom to seek damages or injunctive relief.

Indian law recognises the rights of personality that cover the protection of names, images, voices, or any other identifiable features of a person from unauthorised commercial exploitation. The cases of Amitabh Bachchan and Anil Kapoor have resulted in the rise of strong protections for the unauthorised use of celebrity personas, which also includes AI-generated content.[61] However, these personality-rights cases largely concern unauthorised commercial exploitation of celebrity personas, such as advertisements or merchandise. Non-consensual intimate deepfakes created to harass or silence women have no commercial element and therefore fall outside the core rationale of existing personality-rights doctrine, leaving ordinary women with weaker civil protection than famous public figures. However, the personality rights doctrine in India is still largely concentrated in the area of commercial misappropriation and only to a small extent. The extension of these protections to non-commercial damages, like deepfake pornography for which the purpose is harassment and not profit, entails doctrinal development.

In sum, Indian tort law offers only partial and uneven protection against deepfake abuse. Defamation struggles with the synthetic, non-propositional nature of deepfake pornography and the practical obstacles of anonymity. Privacy is constitutionally recognised but not yet

[60] Akanksha Dubey and Ishita Tripathy, 'DEEPFAKES AND DEFAMATION: A LARGE PERSPECTIVE ON SYNTHETIC HARM' (2025) 4(4) Journal of Legal Research and Juridical Sciences <https://jlrjs.com/wp-content/uploads/2025/08/98.-Akanksha-Dubey.pdf> accessed 10 November 2025

[61] Yash Bhatia, 'BOLLYWOOD VS DEEPFAKES: THE RISE OF PERSONALITY RIGHTS IN INDIA' (*Impact*, 23 September 2025) <https://www.impactonnet.com/impact-stories/bollywood-vs-deepfakes-the-rise-of-personality-rights-in-india-11929.html> accessed 14 November 2025

crystallised into a clear tort of image-based abuse, and personality-rights doctrines are skewed towards commercially exploited celebrity cases. As a result, civil law does not yet provide an accessible, precise remedy for non-consensual intimate deepfakes.

Doctrinal Gaps: Criminal Law: Many different criminal provisions could be analysed to assert liability on deepfakes, but, on top of their piecemeal nature, they were made long before the AI era, and this makes it hard to enforce them. The main doctrinal problem is mens rea. Provisions such as Section 66D IT Act (cheating by personation) and BNS Section 336 (cheating by personation) presuppose an intention to deceive for gain or to cause wrongful loss. By contrast, the primary motive behind many intimate deepfakes is humiliation, intimidation or revenge rather than economic gain, creating uncertainty over whether these offences are conceptually appropriate. For example, to prove under Section 66D that the intention was to deceive or under defamation provisions that the intention was to injure the reputation, causes evidentiary difficulties that, in deepfake cases, can be very hard to overcome.

Figuring out who made the deepfake is a tremendously hard task for the police because digital platforms provide anonymity, and the content can be shared in different countries. In this way, offenders can hide their tracks by using VPNs, encrypted communications, and anonymous accounts. The absence of laws that explicitly prohibit the creation of deepfakes and the spread thereof as separate crimes leads to the situation where there are gaps. This attribution problem is exacerbated by the absence of any deepfake-specific offence that would justify dedicated investigative tools or fast-track procedures. Police are often left to apply generic cybercrime provisions with limited training and without specialised forensic support for synthetic media, leading to under-reporting and low conviction rates.

Most significantly, Indian criminal law contains no explicit provision defining and criminalising non-consensual intimate deepfakes as a standalone offence. Prosecutors must rely on a patchwork of cheating, obscenity, voyeurism, stalking, and privacy-violation provisions, none of which clearly capture the specific wrong of non-consensual synthetic sexualisation. This lack of clear statutory recognition contributes to inconsistent charging practices and weak signalling that such conduct constitutes serious gender-based violence.

Evidentiary Gaps: Deepfakes shake the very foundation of evidence assumptions, whereby the latter audiovisual content is supposed to be a depiction of reality. In fact, traditional verification methods, which are based on the scrutiny of metadata, chain of custody, and witness testimonies, become invalid when a sophisticated AI can produce a convincing fabrication.[62]

Section 63 of The Bharatiya Sakshya Adhiniyam, 2023, is the provision dealing with the admissibility of electronic evidence. It requires that the conditions of Section 63 be met and a certificate under the same provision be produced identifying the electronic record and describing its production.[63] Nevertheless, these requirements were intended for ordinary electronic records and not for AI-generated synthetic media, where there is no 'original'. This framework assumes that there is an 'original' electronic record, such as a CCTV clip or a phone recording, from which copies are derived. On the other hand, deepfakes are completely new creations by AI models and thus may not be based on any original event, which begs entirely new questions about what is being authenticated and how courts should evaluate reliability.

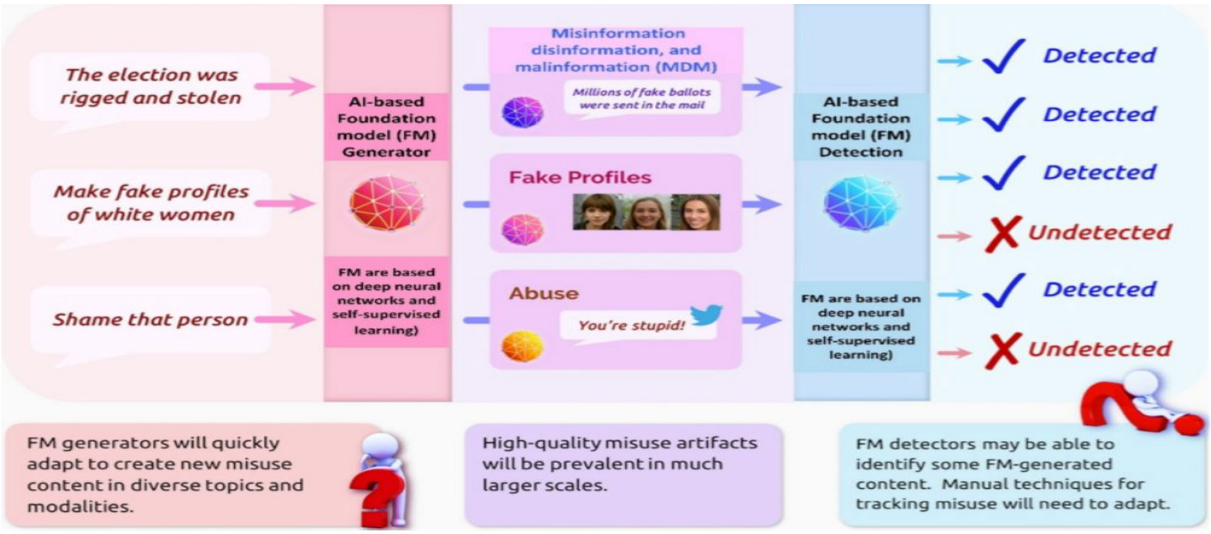
Deepfakes result in the failure of authentication from both sides: firstly, instances where fabricated content can be mistaken for genuine ones; secondly, cases where authentic evidence is rejected due to allegations of deepfakes. Technological tools for deepfake detection are still not perfect, costly, and can be tricked by adversarial methods. At present, there is no foolproof way to verify the authenticity of audiovisual materials. AI-based detection tools can assist courts by estimating the likelihood that a given video is synthetic, but current tools exhibit non-trivial false positive and false negative rates. Evidence law has not yet articulated standards on the admissibility and weight of such detection reports, nor clarified which party bears the burden of proving that content is authentic or a deepfake when this is contested.

Some commentators suggest burden-shifting models, whereby the side disputing the evidence as AI-manipulated must provide enough proof to justify a higher level of

[62] Safee-Naaz Siddiqi et al., 'AI, deepfakes and the burden of proof for digital evidence in litigation' (*Cliffe Dekker Hofmeyr,* 04 November 2025) <https://www.cliffedekkerhofmeyr.com/en/news/publications/2025/Practice/Knowledge-Management/combined-dispute-resolution-and-knowledge-management-alert-4-november-ai-deepfakes-and-the-burden-of-proof-for-digital-evidence-in-litigation> accessed 13 November 2025

[63] Bharatiya Sakshya Adhiniyam 2023, s 63

authentication. The compulsory disclosure of the involvement of AI in the creation of content, the introduction of stricter evidentiary standards for audiovisual material, and the setting up of expert registries with accredited professionals could be among the steps for making the evidence more trustworthy.[64] These uncertainties create serious risks for both victims and accused persons: genuine recordings of abuse may be dismissed as deepfakes, depriving victims of proof, while fabricated exculpatory recordings might be presented as real. Without clear evidentiary rules and judicial training on synthetic media, the criminal justice system may be unable to reliably distinguish truth from fabrication in deepfake cases.



6 Generic workflow of the model to manipulate and generate harmful content taken from (Bommasani et al. 2022)

The recent police activities in Assam, for example, when the officers located a person who made a fake social media profile and to whom they uploaded AI-manipulated intimate images of an influencer by using IP-address logs and device forensics, show the bright side as well as the restrictions of such technical evidence in identifying the creator of deepfake-style content.[65] Addressing these evidentiary challenges will require reforms such as: specialised standards for authenticating synthetic media; accreditation of independent forensic laboratories for deepfake detection; guidelines on admissibility and probative value

[64] Frank Young, 'A Deepfake Evidentiary Rule (Just in Case)' (*University of Illinois Chicago,* 03 July 2025) <https://library.law.uic.edu/news-stories/a-deepfake-evidentiary-rule-just-in-case/> accessed 15 November 2025

[65] Biswa Kalyan Purkayastha, 'Assam man arrested for creating fake profile, AI-generated images of influencer Archita Phukan' *Hindustan Times* (13 July 2025) <https://www.hindustantimes.com/india-news/assam-man-arrested-for-creating-fake-profile-ai-generated-images-of-influencer-archita-phukan-101752383365208.html> accessed 28 November 2025

of AI detection reports; and focused judicial training on evaluating digital evidence in an era of generative AI.

COMPARATIVE GLOBAL PERSPECTIVES

Category	European Union	China	United Kingdom	Australia
Main Laws	EU AI Act 2024	Provisions on Governance of Online Information Content; Deep Synthesis Regulations	Online Safety Act 2023; Data (Use and Access) Act 2025	eSafety Commissioner Framework
Regulatory Approach	Risk-based AI regulation	State-centric information control	Criminalisation of intimate-image deepfakes	Complaints-based takedown model
Deepfake Treatment	Classified as high-risk AI; strict transparency	Allowed only with disclosure; banned for national security/history distortion	Creating, sharing, or requesting intimate deepfakes is a criminal offence	Treated as harmful online content requiring quick removal
Key Requirements	Mandatory labelling, transparency, documentation, risk controls	Mandatory disclosure, platform monitoring, and real-name verification	Criminal penalties; platform removal duties	Mandatory takedowns; the regulator can investigate or escalate to the courts

Enforcement	National Supervisory Authorities	Cyberspace Administration of China (CAC)	OFCOM and law enforcement	eSafety Commissioner (statutory powers)
Policy Philosophy	Protect autonomy, dignity, and democratic processes	Maintain social stability and state control	Protect victims of intimate-image abuse	Fast harm-removal, user-protection focus

Table 2: Comparison between different jurisdictions[66]

Lessons for India: For India, the EU model illustrates the value of regulating high-risk AI systems ex ante rather than relying solely on ex post content moderation. By imposing documentation, transparency and oversight obligations on developers of generative AI, the AI Act treats deepfake risk as a design and deployment problem rather than only a user-behaviour issue. A similar risk-based approach could complement Indian criminal and intermediary liability rules by placing clear responsibilities on AI tool providers to prevent foreseeable misuse for non-consensual intimate deepfakes. The UK's harm-based model shows how platforms can be assigned a statutory 'duty of care' towards users affected by image-based abuse, including deepfake pornography. Instead of depending only on safe harbour doctrines, this method entails carrying out a proactive risk assessment and taking steps to avoid the risks, supported by regulatory enforcement and fines. In the case of India, a selective import of a duty of care for intermediaries regarding nonconsensual intimate deepfakes, together with clear takedown timelines, could greatly enhance the level of protection of victims without the need for a general monitoring obligation for all content. Australia's experience demonstrates the effectiveness of a specialised, independent regulator

[66] 'China:Provisions on the Administration of Deep Synthesis of Internet Information Services' (*FACIA*, 17 September 2025) <https://facia.ai/knowledgebase/chinaprovisions-on-the-administration-of-deep-synthesis-of-internet-information-services/> accessed 20 November 2025; Aislinn O'Connell, 'New Protections for Private and Intimate Images and Videos Online Safety Bill' (*Royal Holloway University of London*, 08 August 2023) <https://www.royalholloway.ac.uk/research-and-education/departments-and-schools/law-and-criminology/news/new-protections-for-private-and-intimate-images-and-videos-in-online-safety-bill/> accessed 25 November 2025

with strong takedown powers for image-based abuse. The eSafety Commissioner has the authority to quickly issue removal notices and can work directly with the platforms, thereby lessening the burden on individual victims of having to chase separate complaints. A comparable Indian institution, whether within MeitY or as an independent online-safety authority, could centralise deepfake complaints, standardise responses, and provide a single point of contact for victims. China's state-centric, heavily censorial model, which combines AI regulation with extensive content control and surveillance, is not compatible with India's constitutional commitments to free speech and democratic pluralism. Its primary comparative value for India lies in highlighting what to avoid: over-broad state discretion, opaque takedown practices, and the conflation of legitimate political speech with harmful synthetic media.

Comparative analysis thus reveals three broad regulatory models: a risk-based ex ante AI-governance model (EU), a harm-based ex post online-safety model with clear duties of care (UK and Australia), and a state-centric surveillance model (China). For India, the most normatively and constitutionally appropriate path is a hybrid of the first two: combining EU-style obligations on AI-tool providers to document, label and mitigate deepfake risks with UK/Australian-style duties of care and rapid takedown mechanisms for platforms, all within the constraints of Articles 14, 15, 19 and 21. Such a hybrid model would move India beyond fragmented, reactive responses and towards a coherent, AI-aware framework for addressing gendered deepfake abuse.

CONCLUSION AND RECOMMENDATIONS

Deepfake abuse in India has emerged as a structurally gendered crisis rather than a marginal cybercrime problem. The analysis in this paper demonstrates that non-consensual intimate deepfakes function as technology-enabled tools of sexualised violence, reputational destruction, and silencing, disproportionately targeting women and those who challenge entrenched hierarchies. These harms are amplified by intersecting axes of caste, religion, class and sexuality, and they produce chilling effects on women's participation in public discourse and democratic life.

At the legal level, India's current response is characterised by three interlocking vacuums. First, there is a doctrinal gap: no provision in the Bharatiya Nyaya Sanhita or Information

Technology Act explicitly defines and criminalises non-consensual intimate deepfakes as a distinct form of image-based sexual abuse, forcing prosecutors to stretch pre-AI offences on cheating, obscenity, voyeurism, privacy violations and harassment beyond their original design. Second, there is a liability gap across the deepfake ecosystem: the respective responsibilities of creators, uploaders, platforms and AI-tool providers are fragmented and unclear, with safe-harbour regimes and draft regulations not yet translating into concrete duties of care or rapid, victim-centred takedown obligations.

Third, there is an evidentiary and procedural gap: evidence law provides no robust standards for authenticating synthetic media or for using deepfake-detection tools in court, and existing civil and criminal procedures are too slow and resource-intensive for victims whose harms are immediate and viral.

Against this backdrop, the three research questions can now be answered directly. Regarding RQ1, the existing Indian legal frameworks in tort, criminal and evidentiary law are not sufficient to address non-consensual intimate deepfakes or to provide effective remedies for women subjected to gendered online harassment. Criminal provisions under the IT Act, BNS and POCSO only partially cover synthetic media and are anchored in concepts such as economic cheating, generic obscenity, and traditional CSAM that do not fully capture the wrong of non-consensual synthetic sexualisation of adults. Tort law and personality-rights doctrines remain underdeveloped for non-commercial, gendered harms, and victims face significant practical barriers in pursuing slow, expensive civil claims against often-anonymous perpetrators. Evidence law, in turn, has not yet caught up with the reality of generative AI, leaving courts without clear tools to distinguish authentic digital recordings from sophisticated deepfakes.

Regarding RQ2, we believe that certain doctrinal and regulatory reforms are required and are strongly supported by Articles 14, 15, 19(2), and 21 of the Constitution. At a minimum, this requires inserting a specific offence into the BNS that defines and criminalises non-consensual intimate deepfakes as a standalone crime, with aggravated forms where the victim is a child, journalist, activist or member of a vulnerable group. It also requires clarifying intermediary duties under the IT Act and the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) framework to ensure rapid takedown, confidential reporting, and victim-centred grievance redressal for intimate

deepfake cases. Further, the Bharatiya Sakshya Adhiniyam must be amended or interpreted to develop authentication standards for synthetic media, rules on admissibility and probative value of deepfake-detection evidence, and guidance on burden-of-proof allocation in disputes about authenticity. These reforms operationalise constitutional guarantees of equality, non-discrimination, dignity and privacy in the context of AI-mediated harms.

In relation to RQ3, the comparative and policy analysis supports a hybrid regulatory model for India that combines ex ante AI governance with ex post online-safety obligations. Ex ante, India should impose obligations on developers and providers of generative AI systems modelled on elements of the EU AI Act and DPDP Rules 2025, requiring risk assessments, technical and organisational safeguards, traceable watermarking or labelling of synthetic content, and monitoring of abuse patterns. Ex post, India should adapt features of the UK and Australian online-safety regimes by establishing a narrowly tailored duty of care on intermediaries for non-consensual intimate deepfakes, including specific takedown timelines, strengthened notice-and-action procedures, and meaningful sanctions for systemic non-compliance, implemented in a manner compatible with Article 19(1)(a) and the permitted restrictions under Article 19(2). A specialised regulatory or coordinating body, housed within or alongside existing cyber law institutions, could centralise deepfake complaints and support consistent enforcement.

KEY RECOMMENDATIONS

Based on Existing Frameworks:

Existing Frameworks	Identified Gaps	Proposed Reforms
Bharatiya Nyaya Sanhita (BNS), IT Act, POCSO, POSH Act	Fragmented laws; no specific clauses on deepfakes; legal uncertainty on AI-generated content; weak penalties for vulnerable groups	Amend BNS to criminalise non-consensual intimate deepfakes; define deepfakes and synthetic media; increase punishments for offences targeting women, children, and disadvantaged groups

IT Intermediary Rules 2025 Draft; SSMI Obligations	Weak enforcement of mandatory labelling; unclear or insufficient intermediary responsibilities	Finalise and enforce IT Rules with mandatory labelling of synthetic content; strengthen safe-harbour obligations for faster removal of deepfake content
Digital Personal Data Protection (DPDP) Rules 2025	Limited algorithmic accountability relating to synthetic media and AI systems	Expand DPDP Rules to require algorithmic impact assessments, bias testing, and transparency disclosures for AI systems handling personal or synthetic data
Bhartiya Sakshya Adhiniyam (Evidence Act)	No authentication standards for AI-generated media; traditional evidence norms are inadequate	Amend evidence law to incorporate deepfake-specific authentication standards; set up certified deepfake labs and expert panels; train judges and law-enforcement personnel
Victim-Support Literature includes StopNCII.org	Poor victim confidentiality; no trauma-sensitive processes; limited support mechanisms	Provide expedited trials with confidentiality; adopt trauma-informed procedures; use hash-sharing tools; offer free legal assistance, mental-health support, and helplines
Digital Literacy and Online Safety Research	Limited public awareness and digital literacy on deepfake risks	Mandate digital-literacy education nationwide; launch awareness campaigns on deepfake harms, legal remedies, and reporting pathways

International Regulatory Cooperation Models & Treaties	Weak cross-border enforcement; no international standards for deepfake regulation	Strengthen MLATs; join international regulatory collaborations; engage in multilateral forums to create global standards for deepfake regulation

Reform Roadmap for Addressing Deepfake Abuse in India:

Focus	Key Measures Proposed	Purpose	Rationale
Legislative Reform	Criminal and Civil Law Framework	• Introduce a deepfake-specific offence under the Bharatiya Nyaya Sanhita (BNS) • Refine and align relevant provisions of the Information Technology Act • Consider a statutory tort for non-consensual intimate imagery	To close existing legal gaps, clearly criminalise deepfake abuse, and provide victims with accessible civil remedies rather than forcing them into doctrinal contortions
Regulatory Reform	Platform and Data Governance	• Finalise and strengthen IT Intermediary Rules 2025 • Finalise DPDP Rules 2025 • Impose concrete duties of care, transparency obligations, and rapid takedown requirements on platforms and AI generator services	To shift responsibility from victims to intermediaries and service providers, and ensure accountability across the deepfake ecosystem

		• Explicitly recognise gendered harms	
Evidentiary Reform	Proof and Procedure	• Update the Bharatiya Sakshya Adhiniyam and allied procedural rules • Establish standards for the authentication of synthetic media • Regulate the admissibility and use of deepfake-detection tools • Address risks of false positives and false negatives	To equip courts with reliable tools to assess synthetic evidence without undermining due process or evidentiary integrity
Institutional Reform	Capacity Building and Victim Support	• Develop specialised expertise within cyber-crime units, regulators, and the judiciary • Establish dedicated victim-support mechanisms offering legal, psychological, and technical assistance tailored to deepfake cases	To ensure effective enforcement, informed adjudication, and meaningful support for victims navigating complex technological harm