



International Journal of Research Publication and Reviews

Journal Homepage: www.ijrpr.com ISSN 2582-7421

Deepfakes and the Law in India: Addressing Misinformation, Identity Misuse and Emerging Regulatory Challenges

Aanchal Ajit Malpani

UG Student, Cyber Law, Ajeenkya DY Patil University, Pune

ABSTRACT

Generative artificial intelligence has, in a short span, altered how audio-visual content is produced and shared. Deepfakes are audio, images or video that are synthetically produced or manipulated to mimic real people or events convincingly. They raise legal questions spanning misinformation, impersonation, privacy, reputation, fraud, sexual exploitation and personality rights. India has not traditionally dealt with these harms through a deepfake-specific law. It has relied instead on the Information Technology Act, 2000, the rules governing intermediaries, and the general body of criminal and civil law. That position has shifted considerably. Following their 2026 amendment, the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 now deal directly with “synthetically generated information”, and attach duties relating to prevention, labelling, metadata, technical provenance, grievance redressal and takedown. This paper tests whether the framework now taking shape in India is adequate. It considers the IT Act, the 2026 IT Rules, the Bharatiya Nyaya Sanhita, 2023, constitutional privacy doctrine, and recent case law on personality rights and AI-generated material. It then sets India’s approach against the European Union’s Artificial Intelligence Act. Persistent problems are identified in detection, attribution, cross-border enforcement, remedies for victims, platform accountability and the safeguarding of legitimate speech. The paper contends that the most suitable model is one organised around harm, bringing together transparency, provenance, swift remedies, specialised enforcement and constitutional safeguards.

Keywords: Deepfakes; Artificial Intelligence; Cyber Law; Misinformation; Personality Rights; Intermediary Liability.

INTRODUCTION

The arrival of generative artificial intelligence has unsettled the link between digital content and its authenticity. Recordings, photographs and videos can now be fabricated or edited so realistically that an ordinary viewer may be unable to tell whether the person shown ever actually said, did or appeared in what is depicted. For this reason deepfakes are more than a technical novelty. They pose questions of evidence, identity, autonomy, reputation and accountability in online environments.

In common usage, a “deepfake” is synthetic media created or altered with artificial intelligence. Typically someone’s face, voice, appearance or behaviour is fabricated or changed so that the output seems genuine. Nothing about the technology is unlawful in itself. It has legitimate applications in entertainment, teaching, accessibility, art, research and elsewhere. Legal difficulty begins only when synthetic material is deployed to mislead, impersonate, exploit, defraud or otherwise inflict unlawful harm.

The harms that can follow vary widely. Someone may be shown in a fabricated video saying things they never said. A voice clone may be used to trick a person into sending money. Synthetic intimate imagery may be shared without the depicted person’s consent. An altered recording may be circulated to injure someone’s standing. One and the same technical process can therefore engage several legal fields at once: cybercrime, criminal law, privacy, defamation, intellectual property, consumer protection and the regulation of intermediaries.

India has responded through a mix of technology-neutral statutes and platform duties that are becoming progressively more targeted. The Information Technology Act, 2000 (“IT Act”) includes provisions on identity theft, cheating by personation, breaches of privacy, and obscene or sexually explicit electronic content. Offences of cheating, personation, forgery and defamation are found in the Bharatiya Nyaya Sanhita, 2023 (“BNS”). Beyond statute, constitutional case law, above all the recognition of privacy as a fundamental right,

¹ Offers a further normative lens.

Of the recent developments, the most consequential is the 2026 amendment to the

Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (“IT Rules”). The amendment brings in the category of “synthetically generated information” (“SGI”) and places particular duties on intermediaries that enable its creation or spread. Among these are measures

directed at specified categories of unlawful synthetic content, together with labelling, metadata and provenance requirements for SGI that falls outside those categories.

Against this background, the paper asks whether India's developing framework deals adequately with the harms deepfakes cause while still preserving constitutional protection for legitimate expression. Its focus is on how the law treats misinformation and the misuse of identity. It does not proceed on the premise that synthetic media is unlawful *per se*.

THE CHARACTER OF DEEPFAKES AND THEIR LEGAL RELEVANCE

Deepfakes come in several varieties: face swaps, cloned voices, lip-synced footage, synthetic avatars, and audio-visual content generated entirely from scratch. How the law views any of these turns not just on the method of production. It also turns on why the material was made, where and to whom it is shown, and what effects it has.

Authenticity is the first problem. Older forms of manipulated media often gave themselves away through visible flaws, whereas more advanced generative tools can yield output that a typical viewer finds believable. The evidentiary consequence is that digital material can be convincing and yet not be real.

The second problem concerns scale and velocity. Within minutes, a damaging synthetic clip can be duplicated, re-shared, translated and spread across numerous services. Taking down the first upload may leave other copies available. Any remedy offered to a victim must therefore reach not only the original posting but also its onward circulation.

Attribution is the third problem. Those who create such material may operate through anonymous profiles, overseas services, throwaway identifiers or tools that frustrate investigation. A legal ban achieves little in practice if the person responsible cannot be traced or if the relevant evidence is lost.

The 2026 IT Rules tackle part of this difficulty by defining SGI. Under the definition, SGI is audio, visual or audio-visual information that has been artificially or algorithmically created, generated, modified or altered through a computer resource so that it seems real or authentic. It must also portray a person or event in a way that may be indistinguishable from an actual natural person or real-world event. Carve-outs exist for certain routine editing and for accessibility, educational, research and good-faith purposes. The regulatory category is accordingly broader than the everyday notion of a "deepfake", yet it does not prohibit every act of editing.

THE INFORMATION TECHNOLOGY ACT, 2000

A. Identity Theft and Cheating by Personation

Section 66C of the IT Act penalises anyone who, acting fraudulently or dishonestly, makes use of credentials belonging to another person, such as that person's electronic signature or password, or any other identifying marker unique to them. Section 66D covers cheating by ² personation carried out through a communication device or computer resource. Either provision may come into play where a deepfake forms part of a fraudulent scheme to impersonate someone.

Consider a cloned voice that mimics a senior executive and persuades an employee to move funds. Provided the evidence establishes the statutory ingredients, that conduct could amount to cheating by personation. In the same way, where a fabricated video is used to secure some advantage by deceit, general criminal provisions may apply alongside those specific to cyberspace.

These sections have a clear limit. They do not make the creation of a deepfake an offence in its own right. The prosecution has to prove fraudulent or dishonest conduct and, for section 66D, the elements of cheating by personation. The legality of the material is therefore a function of what was done in making or using it, not of the material alone.

B. Privacy and Synthetic Intimate Imagery

Section 66E makes it an offence to intentionally or knowingly capture, publish or transmit an image of a person's private area without consent, in circumstances that violate that ³ person's privacy. Obscene and sexually explicit electronic material is governed by sections 67 and 67A, and section 67B makes special provision for sexually explicit material ⁴ involving children.

In cases of synthetic intimate imagery, these provisions may carry considerable weight. A legal difficulty emerges, though, where the image has been generated from nothing rather than produced by altering a genuine photograph. Whether a provision framed around capturing or publishing an "image" extends to every kind of synthetic depiction may turn on the facts and on how courts read it. Some of this uncertainty is removed by the 2026 IT

Rules, which specifically name non-consensual intimate imagery and privacy-invasive SGI ⁵ among the categories of prohibited synthetic information.

SYNTHETIC MEDIA UNDER THE 2026 IT RULES

The 2026 amendments mark a significant shift. They take Indian intermediary regulation away from a generic treatment of harmful online material and towards a framework designed specifically for synthetic media.

Besides defining SGI, the amended Rules set out a due-diligence regime devoted to it. An intermediary that facilitates the creation, generation, modification, publication, transmission, sharing or dissemination of SGI must put in place reasonable and appropriate technical measures to prevent specified categories of unlawful synthetic content. Those categories comprise child sexual exploitative and abuse material; non-consensual intimate imagery; obscene or sexually explicit content; material that invades privacy; false documents or false electronic records; and certain content that falsely portrays a natural person or real-world event in a manner likely to deceive as to identity, voice, conduct, ⁶ action, statement or occurrence.

A line is also drawn between prohibited SGI and other synthetic material. SGI falling outside the prohibited categories has to carry a prominent label. So far as is technically feasible, the Rules also call for permanent metadata or some other suitable technical provenance mechanism. This includes a unique identifier able to identify the computer resource through which the information was created, generated, modified or altered. Intermediaries may not allow the prescribed label or provenance information to be modified, suppressed or removed.

Additional obligations fall on significant social media intermediaries. These platforms must ask users to state whether content is synthetically generated, use appropriate technical measures to check those statements, and label prominently any content confirmed as ⁷ synthetic. The heavier burden reflects the scale at which large platforms operate and their capacity to run technical verification systems.

The 2026 framework matters for two reasons. It acknowledges, first, that synthetic media may cause harm even when it does not fit comfortably within an established cyber offence. Second, rather than depending solely on punishment once harm has occurred, it adds obligations aimed at prevention and transparency.

SAFE HARBOUR, PLATFORM DUTIES AND FREE SPEECH

Deepfake regulation cannot be separated from the question of when online platforms are answerable for what their users post. Section 79 of the IT Act shields intermediaries from liability for information supplied by third parties, but that protection is not unconditional: the intermediary must satisfy the requirements the statute lays down and must exercise due diligence.

On the constitutional plane, no ruling matters more here than *Shreya Singhal v. Union of India*. There, the Supreme Court invalidated section 66A of the IT Act and examined how intermediary obligations relate to free speech. It also interpreted the “actual knowledge” requirement in section 79(3)(b) as tied to legally valid orders and notifications. Under that ⁸ reading, private parties are not left to decide freely whether content is illegal.

Regulating deepfakes involves a hard trade-off. Damaging synthetic content can travel fast and may call for immediate intervention, especially where intimate imagery, fraud or misuse of identity is involved. Yet filtering tools that run without human judgement are prone to error, and may treat parody, satire, creative works, teaching material or fair criticism as though it were unlawful.

Rather than leaving this balance to chance, the amended Rules specify in advance which kinds of content are covered, and pair that specification with technical safeguards, routes for complaint and obligations of procedure. How they are implemented still matters, however. Should platforms be pushed to take down doubtful material merely to limit their legal exposure, the effect could reach well beyond unlawful deepfakes. If, conversely, they respond too slowly, the victim may sustain reputational or privacy harm that cannot be ⁹ undone.

A sound legal approach thus demands both effective takedown mechanisms and procedural protections. Keeping safety and lawful expression in balance depends on several things: explaining why material was removed, preserving evidence, providing accessible grievance channels, and offering genuine review.

SAFEGUARDING IDENTITY: PRIVACY AND PERSONALITY RIGHTS

Where deepfakes are used to misappropriate someone’s identity, the constitutional guarantee of privacy supplies a useful starting point for analysis. The Supreme Court’s judgment in *Justice K.S. Puttaswamy (Retd.) v. Union of India* recognised privacy as a right protected by the Constitution, anchoring it in Article 21 and treating it as bound up with a person’s liberty, dignity and capacity for self-determination. The Court clarified that the right is not absolute. Any intrusion must meet constitutional tests including legality, a ¹⁰ legitimate state aim and proportionality.

A deepfake can infringe privacy without extracting anything from a confidential store of data. Someone’s face, voice, likeness or identity may be taken and rebuilt into an event that never took place. The injury here is not only a loss of secrecy. It is a loss of control over one’s own identity and of the ability to avoid being falsely portrayed.

Indian courts have also built up civil protection for personality and publicity rights. In *Anil Kapoor v. Simply Life India*, the Delhi High Court dealt with the unauthorised use of the actor’s name, image, voice and other attributes. Rulings of this kind show that established intellectual property and personality-rights doctrines can furnish relief where a person’s ¹¹ identity is exploited commercially or misrepresented.

Of particular relevance to AI-driven misuse is the Delhi High Court’s 2025 ruling in *Hrithik Roshan v. Ashok Kumar/John Doe & Ors*. The case concerned the alleged unauthorised exploitation of the plaintiff’s name, image, voice and personality by means of artificial intelligence, morphing and similar technologies. At the interim stage, the Court extended protection against specified kinds of misuse and ordered the takedown of identified infringing content. The ruling shows that civil injunctions and personality-rights doctrine can be deployed against new technological harms even in the absence of a ¹² dedicated deepfake statute.

DEEPFAKE CONDUCT UNDER THE BHARATIYA NYAYA SANHITA, 2023

The BNS widens the range of criminal-law responses available against conduct involving deepfakes. The offence of cheating by personation sits in section 319, and it reaches anyone who poses as another individual, whether that individual exists or is fictitious. Section 318 addresses cheating more broadly, where deceit causes or is likely to cause harm to body,¹³ mind, reputation or property.

Forgery is dealt with in section 336, whose statutory scheme extends to false electronic¹⁴ records. This could become pertinent where synthetic content is used, with the necessary fraudulent or injurious intent, to fabricate or tamper with an electronic record. Defamation falls under section 356, which reaches imputations conveyed through visible¹⁵ representations.

Taken together, these provisions indicate that Indian criminal law already holds several instruments for dealing with harmful deepfake conduct. Their use, however, hinges on the facts and on the statutory ingredients being met. A deepfake produced purely as satire, for instance, raises quite different legal issues from a fabricated video employed to extract money through deceit.

The core difficulty, then, is not merely whether a suitable provision exists. It is whether investigators and courts can correctly link the technology to what an offence requires: proving intention, preserving dependable digital evidence, identifying who is responsible, and separating unlawful conduct from protected expression.

A COMPARATIVE VIEW: THE EUROPEAN UNION

The European Union's Artificial Intelligence Act provides a helpful point of comparison. Article 50 imposes transparency duties in respect of certain AI-generated or manipulated content. Most relevantly, those who deploy systems that generate or manipulate image, audio or video content amounting to a deepfake must disclose that the content is artificially generated or manipulated. Exceptions apply, among them certain artistic, creative, satirical¹⁶ and fictional contexts.

The EU model shows what transparency can achieve. It does not regard every synthetic depiction as unlawful by nature. Instead, it accepts that users will come across legitimate¹⁷ synthetic content and aims to ensure they can recognise it as artificial.

India's 2026 IT Rules likewise rely on labelling and provenance requirements, but they go¹⁸ further by imposing preventive duties in respect of specified unlawful SGI. The two regimes thus share a focus on transparency while differing in design and reach.

What the comparison suggests is that a mature regulatory framework should not depend on any single remedy. Transparency can lessen deception. Technical provenance can aid attribution and verification. Removing offending material stops the damage from compounding, and the most serious abuses can be pursued through criminal prosecution or civil action. Each of these instruments, though, ought to work in tandem with protections for speech that remains within the law.

GAPS THAT PERSIST

Notwithstanding these regulatory advances, a number of problems remain.

To begin with, India's regulation of deepfakes is still spread across the IT Act, the IT Rules, criminal law, constitutional law and civil remedies. Depending on the harm, a victim may have to locate quite different provisions.

Detection technology, moreover, is imperfect. Generative systems develop rapidly, and detection tools may yield false positives or lose effectiveness as newer generation methods appear.

Attribution also remains a challenge. A damaging deepfake may be produced in one country, uploaded via another and spread worldwide. Gathering evidence and identifying those responsible may depend on international cooperation.

Nor does taking content down necessarily bring the harm to an end. Copies, screenshots, downloads and re-uploads can outlast the disappearance of the original. The legal system accordingly needs effective mechanisms for preservation, tracing and coordinated takedown.

Strong duties on intermediaries also carry a risk of over-removal. Faced with doubtful material, a platform may prefer to take it down rather than decide whether it is lawful. This can affect satire, parody, artistic expression, research and criticism.

Finally, victims need remedies they can actually reach. However technically advanced a legal framework may be, its practical worth is limited if an ordinary victim cannot promptly report harmful content, preserve evidence, secure its removal and pursue judicial relief.

PROPOSALS FOR REFORM

India's emerging framework could be reinforced by an approach that is centred on harm and attentive to rights.

Parliament could, in the first place, examine whether a consolidated statutory framework is needed to sit alongside the technology-neutral provisions of the IT Act and the BNS. Legislation of this kind could bring clarity to definitions, offences, victim remedies, evidentiary standards and platform responsibilities, without criminalising legitimate synthetic media.

Labelling and provenance standards should also work across platforms. An identifier of synthetic content that is lost once the content is downloaded or re-posted offers little real benefit. Standards should therefore support its persistence through reasonable technical means.

In addition, platforms should operate fast-track grievance channels for non-consensual intimate imagery, identity impersonation, financial fraud and serious reputational harm. These channels ought to acknowledge receipt of complaints, follow procedures for preserving evidence, and communicate decisions clearly.

Removal and evidence preservation should, further, be treated as separate matters. When content is taken down on the ground that it allegedly constitutes an offence, the relevant metadata and records should be retained in accordance with law. Investigators can then establish its source, timing and spread.

Regulatory safeguards should also shield legitimate expression. Satire, parody, and artistic and educational material ought not to be treated as unlawful simply because they are synthetic. The legal framework should concentrate on deception, harm, unlawful sexual content, fraud, invasion of privacy and other clearly defined categories.

Police and other enforcement agencies, moreover, should be trained specifically to detect synthetic media and to handle digital forensics, electronic evidence and the chain of custody. Awareness among judges and prosecutors matters no less. The legal question frequently turns on proving what was generated, by whom, when and with what intent.

Lastly, India should deepen international cooperation on cross-border evidence, requests to platforms, attribution and enforcement. Because the spread of deepfakes is transnational by nature, domestic remedies may not always be enough on their own.

CONCLUSION

Deepfakes call into question a foundational premise of digital communication: that an audio-visual record can be trusted as evidence of who someone is or what happened. Misused, they can damage privacy, reputation, personality rights, financial security and personal autonomy. Synthetic media nonetheless has legitimate applications, which makes a blanket ban difficult to justify in law and to achieve in technical terms.

India's legal response has advanced considerably. Through the IT Act, the law reaches identity theft and personation, privacy violations, obscene and sexually explicit material, and the liability of intermediaries. The BNS supplies general criminal offences covering cheating, personation, forgery, false electronic records and defamation. Constitutional privacy jurisprudence contributes a rights-based framework. Recent Delhi High Court rulings also demonstrate that personality-rights remedies can answer AI-enabled misuse of identity.

Among these developments the 2026 IT Rules stand out, since they explicitly recognise synthetically generated information. They attach duties covering what synthetic content is forbidden, how it must be labelled, what metadata and provenance it must carry, how declarations are verified, and what due diligence intermediaries owe. The debate thereby moves beyond the blunt question of whether deepfakes are lawful to a sharper enquiry: which kinds of synthetic content are harmful, who is obliged to act, and what safeguards ought to apply.

What remains is the task of building a framework that works without overreaching. Regulation should accordingly bring together prevention, transparency, provenance, swift remedies for victims, preservation of evidence, specialised enforcement and constitutional protection for lawful expression. An approach of this kind can address the particular harms deepfakes cause while acknowledging that the technology is not, in itself, the same thing as illegality.

Ultimately, the course of deepfake regulation in India will turn not just on the creation of new offences but also on whether existing and emerging rules can be applied promptly, uniformly and in proportion to the harm. That is the challenge in an environment where fabricated material can spread faster than conventional legal remedies are able to react.

FOOTNOTES — BLUEBOOK 20TH EDITION

1. Information Technology Act, No. 21 of 2000, §§ 66C–66E, 67, 67A, 67B, 79, India Code.
2. *Id.* §§ 66C–66D.
3. *Id.* § 66E.
4. *Id.* §§ 67, 67A, 67B.
5. Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, r.2(1)(wa), as amended by G.S.R. 120(E) (Feb. 10, 2026).
6. *Id.* r. 3(3).
7. *Id.* r. 4(1A).

8. Shreya Singhal v. Union of India, (2015) 5 SCC 1 (India).
9. Id.
10. Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1 (India).
11. Anil Kapoor v. Simply Life India, 2023 SCC OnLine Del 6914 (India).
12. Hrithik Roshan v. Ashok Kumar/John Doe & Ors., CS(COMM) 1107/2025 (Delhi High Court Oct. 15, 2025).
13. Bharatiya Nyaya Sanhita, No. 45 of 2023, §§ 318–319 (India).
14. Id. § 336.
15. Id. § 356.
16. Regulation (EU) 2024/1689 of the European Parliament and of the Council of June 13, 2024, art.50, 2024 O.J. (L 1689).
17. Id. art. 50(4).
18. Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, r.3(3), as amended by G.S.R. 120(E) (Feb. 10, 2026).