*Oxford Law Review*, Vol. 1, Issue 1, 2026 DOI: https://10.5281/zenodo.23540582

# Synthetic Deception and Criminal Responsibility: Are India's Existing Laws Adequate to Address AI-Generated Deepfakes?

**Shifali Singh**[1]

© The Author(s), under exclusive licence to Oxford Law Review 2026.

AI-generated deepfakes expose a difficult question for criminal law: when synthetic media convincingly depicts a real person saying or doing something that never occurred, is the wrong adequately captured by offences drafted around impersonation, obscenity, privacy, forgery, cheating and defamation? This study argues that India does not face a complete legal vacuum. The Information Technology Act 2000, the Bharatiya Nyaya Sanhita 2023 (BNS), the Protection of Children from Sexual Offences Act 2012 (POCSO), and the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021, as amended in February 2026, already address significant parts of the problem. The central weakness is instead one of fit, coherence and enforcement: liability depends on the harm, the actor's mental state, the kind of content, and the precise act of creation, possession, publication or use. Some provisions fit a deepfake only where additional statutory elements are proved; likeness alone does not automatically establish identity theft, forgery, voyeurism or defamation. The 2026 Rules add synthetic-media-specific duties for intermediaries, including prevention of unlawful synthetic information and labelling and provenance requirements for permissible material. Yet platform duties do not themselves create a comprehensive, general offence of harmful synthetic impersonation, and removal is not a substitute for criminal accountability or victim redress. The study proposes a harm-based, technology-neutral response: enforce existing offences accurately; issue clear prosecution guidance; preserve digital evidence; provide fast, reviewable remedies; and consider a narrowly drafted offence for intentional, materially deceptive impersonation causing specified serious harm, with safeguards for satire, art, journalism, parody, consent and public-interest expression.

A deepfake is commonly understood as audio, image or video generated or materially altered with artificial-intelligence techniques so that it appears to depict a real person or event. The term covers very different conduct: a consensual film effect, political satire, a fabricated endorsement, a voice-cloned fraud call, a non-consensual sexual image, and a forged video

[1] **Chandigarh Law College, CGC University, Mohali, Punjab, India; ✉ s.shifali.2006@gmail.com**



Copyright © Oxford Law Review. (www.oxfordlr.com)

*Oxford Law Review*, Vol. 1, Issue 1, 2026 DOI: https://10.5281/zenodo.23540582

circulated to provoke communal tension are not morally or legally equivalent. The technology is a means; the legally relevant questions concern what the content represents, how it is used, the harm it causes or risks, and what the creator or distributor intended or knew. The legal debate is often framed as a binary: either existing law is sufficient or a new deepfake offence is essential. That framing is too blunt. Criminal law is offence-specific. A synthetic video may amount to cheating by personation if used dishonestly to induce delivery of property; it may support a forgery charge if the statutory requirements concerning a false document or electronic record and the relevant intent are satisfied; it may fall within provisions on obscene or sexually explicit electronic material; or it may constitute defamation, intimidation or another offence depending on its content and use. The same video, however, may not satisfy every one of those offences. A false depiction is not automatically a forged electronic record for every statutory purpose, and a face swap is not automatically identity theft under section 66C of the Information Technology Act 2000.

This study asks: do India's criminal and information-technology laws provide a coherent and effective framework for addressing the distinct harms caused by AI-generated deepfakes? It answers that the framework is substantial but incomplete. The law is strongest when the synthetic content is used as an instrument of an already recognized wrong, such as fraud, sexual exploitation, extortion or deliberate defamation. It is less certain where the harm lies in the non-consensual fabrication of a person's likeness itself, where serious reputational or dignitary injury occurs without a clear fit with an existing offence, or where attribution and proof are difficult. The February 2026 amendments to the IT Rules are important because they expressly regulate synthetically generated information, but their principal role is intermediary due diligence and content governance, not the replacement of offence-by-offence criminal analysis. The argument proceeds in five stages. It first defines the problem and its relevant distinctions. It then maps the main statutory provisions, considers the 2026 IT Rules, analyses case law and practical enforcement, and identifies the limits of the present framework. It concludes with targeted reforms that avoid both technological exceptionalism and overcriminalization.

"Deepfake" is a technological label, not a legal element. Courts should therefore avoid treating the use of AI as proof of criminality. A useful starting point is to distinguish four dimensions.

First, distinguish fabrication from falsity. Some synthetic content is fictional, obviously stylized or labelled. Some is realistic but used in a context where viewers understand it to be parody. Other content is designed to make a reasonable recipient believe that a person actually said or did something. Deceptive presentation, not merely synthetic production, is often the key feature.

Second, distinguish creation, possession, publication and use. A person who creates a private synthetic clip, a person who posts it publicly, a person who forwards it while believing it genuine, and a person who uses it to extract money may occupy different positions. A sound criminal analysis identifies each actor's conduct and mental state rather than treating every link in the circulation chain as equally culpable. Sharing may be evidence of intent, but it does not invariably prove knowledge of falsity or the intent required by a particular offence.


Copyright © Oxford Law Review. (www.oxfordlr.com)

*Oxford Law Review*, Vol. 1, Issue 1, 2026 DOI: https://10.5281/zenodo.23540582

Third, distinguish harms. Financial fraud, sexual humiliation, reputational injury, coercion, political deception, identity misuse and threats to public order engage different legal interests. A single general prohibition on "deepfakes" risks both overbreadth and under-inclusion: it could criminalize harmless fiction while failing to describe precisely the conduct that makes a synthetic image harmful.

Fourth, distinguish authenticity from attribution. A forensic expert may conclude that a video was manipulated without identifying who manipulated it. Conversely, account records may identify an uploader without proving that the uploader created the file or knew it was false. Both questions matter, and neither should be collapsed into the other.

The most useful mental model is therefore a "conduct–mental state–harm–proof" matrix. For each incident, ask: what did the accused do; what did the accused intend, know or have reason to believe; what protected interest was harmed or put at legally recognized risk; and what admissible evidence connects the accused to each element? This model reduces the temptation to infer guilt from the mere presence of AI.

The IT Act is the central technology-specific statute. Section 66C criminalizes fraudulent or dishonest use of another person's electronic signature, password or other unique identification feature. Section 66D criminalizes cheating by personation through a communication device or computer resource. Section 66E addresses intentional or knowing capture, publication or transmission of an image of a person's private area without consent in circumstances violating privacy. Sections 67 and 67A address publication or transmission in electronic form of obscene material and material containing sexually explicit acts, respectively; section 67B deals with specified child-sexual material-related conduct. Section 69A provides a statutory route for blocking public access to information, while section 79 establishes conditional intermediary protection subject to statutory conditions and due diligence.[2]

These provisions are significant but should not be overstated. Section 66C is not a universal prohibition on using someone's face or voice. Its language concerns an electronic signature, password or other unique identification feature; whether a particular digital identifier falls within the provision requires analysis of the statutory definition and facts. Section 66D requires cheating by personation. A fabricated voice used to persuade a victim to transfer funds may fit the provision if the ingredients of cheating and personation are established; a non-commercial parody ordinarily does not become a section 66D offence merely because it imitates a public figure.

Section 66E is also not a general image-based abuse provision covering every false sexual depiction. Its text refers to an image of a person's private area and specified privacy circumstances. A synthetic image may not involve the actual capture of the depicted person's private area. Depending on the facts, sections 67 or 67A, other statutes or general criminal provisions may be more relevant, but their own elements must be satisfied. The difference

[2] Information Technology Act 2000, ss 66C–66E, 67, 67A, 67B, 69A and 79.


Copyright © Oxford Law Review. (www.oxfordlr.com)

*Oxford Law Review*, Vol. 1, Issue 1, 2026 DOI: https://10.5281/zenodo.23540582

between actual intimate imagery and a synthetic fabrication is legally important, even where the victim's distress is comparable.

Sections 67 and 67A may apply to qualifying synthetic material when it is published or transmitted electronically. But they are not general offences of falsity or non-consensual likeness use. Their application turns on the statutory tests concerning obscenity or sexually explicit acts and on the act of publication or transmission. Section 67B and POCSO require particular care where synthetic material depicts a child: the wording and judicial interpretation of the relevant provisions must be examined rather than assuming that every fictional or AI-generated image is covered in exactly the same way as a depiction involving an actual child.

The BNS, in force from 1 July 2024, supplies general criminal offences that may apply regardless of whether the medium is AI-generated. Section 318 concerns cheating; section 319 concerns cheating by personation. Section 336 addresses forgery, with related provisions dealing with forged documents or electronic records and their use as genuine. Section 351 concerns criminal intimidation; section 353 addresses statements conducing to public mischief in the circumstances specified by the provision; and section 356 concerns defamation. The BNS also includes gender-specific offences such as voyeurism and stalking, but the fit of those provisions to a wholly synthetic depiction is not automatic.[3]

A fake video used to obtain a loan, induce a transfer, or make a victim surrender property may support cheating charges if deception, dishonest inducement and the remaining statutory elements are proved. A fabricated electronic record may raise forgery questions where it falls within the statutory concepts of a false document or electronic record and the required intent exists. The prosecution should not assume that every misleading video is "forgery" simply because it is false. The legal analysis must connect the medium to the definition and purpose specified in the statute.

Defamation may be relevant when a deepfake communicates an imputation concerning a person with the elements required by section 356, subject to its exceptions. But a false video does not by itself establish every element of criminal defamation. Context, imputation, intention or knowledge, publication and applicable exceptions must be assessed. Similarly, section 353 should not be treated as a catch-all misinformation offence: the prosecution must establish the particular type of statement and the public-mischief-related conditions contained in its text. Broadly labelling content "fake news" cannot substitute for proving an offence.

The BNS's technology-neutral offences are an advantage because offenders cannot necessarily escape liability merely by using a new tool. Yet technology neutrality is not the same as perfect coverage. General offences were designed around particular harms and mental states; they may not expressly capture the non-consensual appropriation of likeness as a distinct injury. The correct question is not whether a deepfake is novel, but whether a specific offence's elements fit the conduct without strained interpretation.

[3] Bharatiya Nyaya Sanhita 2023, ss 318, 319, 336, 340, 351, 353 and 356.


Copyright © Oxford Law Review. (www.oxfordlr.com)

*Oxford Law Review*, Vol. 1, Issue 1, 2026 DOI: https://10.5281/zenodo.23540582

The Protection of Children from Sexual Offences Act 2012 is especially relevant when a deepfake depicts a child in sexualized circumstances. Section 11 includes forms of sexual harassment involving a real or fabricated depiction of a child in certain circumstances. Sections 13–15 address the use of children for pornographic purposes and specified conduct involving pornographic material. The IT Act's section 67B is also relevant to child-sexual material.[4] Nevertheless, synthetic-only depictions raise difficult interpretive questions: does a provision require a real child to have been used or involved, or can a wholly artificial depiction fall within the statutory language? The answer depends on the particular provision and authoritative interpretation; it should not be asserted categorically without direct authority. In Just Rights for Children Alliance v S Harish, the Supreme Court addressed the statutory scheme concerning child sexual exploitation material and the interpretation of the IT Act and POCSO. The decision is important to understanding the protective purpose and scope of those provisions, but it should not be cited as a blanket holding that every AI-generated fictional depiction is covered. The principled response is to protect actual children robustly while clarifying any genuine statutory ambiguity through legislation rather than stretching penal language by analogy.

Deepfake prosecutions remain subject to ordinary requirements of actus reus, mens rea, causation and proof beyond reasonable doubt. The availability of an AI tool does not establish that the accused generated the content, knew it was false, intended to deceive, or caused the alleged harm. Where an offence requires dishonest intention, sexual intent, knowledge, or intent to harm reputation, those requirements remain essential. Criminal law's legality principle also matters: courts should interpret existing provisions purposively but should not create a new offence by analogy merely because conduct is harmful or technologically novel.

The regulatory position changed materially in February 2026. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021, as amended by the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules 2026, introduce an express definition of "synthetically generated information" for relevant audio, visual and audio-visual material. The definition focuses on material artificially or algorithmically created, generated, modified or altered so that it appears real or authentic and depicts an individual or event in a manner likely to be perceived as indistinguishable from a natural person or real-world event, subject to specified exclusions.[5]

The amended Rule 3(3) requires intermediaries offering tools or services that enable synthetic information to deploy reasonable and appropriate technical measures to prevent users from creating, generating, modifying, publishing or disseminating synthetic information that violates applicable law. The listed categories include child sexual exploitative and abuse material, non-consensual intimate imagery, obscene or sexually explicit material, false documents or

[4] Protection of Children from Sexual Offences Act 2012, ss 11, 13–15.

[5] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021, rr 3 and 4, as amended by the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules 2026, Gazette of India, Extraordinary, GSR 120(E), 10 February 2026, as corrected by GSR 148(E), 26 February 2026.


Copyright © Oxford Law Review. (www.oxfordlr.com)

*Oxford Law Review*, Vol. 1, Issue 1, 2026 DOI: https://10.5281/zenodo.23540582

electronic records, and synthetic depictions that misrepresent a person's identity, voice, conduct, action or statement, or an event, in a manner likely to deceive. Synthetic information not covered by the prohibition must, subject to the rule's terms and technical feasibility, be prominently labelled and include permanent metadata or another provenance mechanism, including a unique identifier. Significant social media intermediaries have additional obligations concerning user declarations, verification mechanisms and labels.

The Rules also strengthen removal obligations. Under the amended Rule 3(1)(d), where the intermediary receives actual knowledge through the specified legal routes, unlawful information must be removed or access disabled within three hours. The Rules provide specialized grievance timelines and duties concerning certain categories of harmful content. These requirements are legally important and should be acknowledged in any current assessment; it would be inaccurate, as of October 2026, to describe India as having no synthetic-media-specific intermediary regulation.

Three qualifications are essential. First, the Rules primarily impose due-diligence duties on intermediaries. They do not automatically make every synthetic depiction a criminal offence, nor do they dispense with the need to identify an underlying illegality. Secondly, a platform's failure to comply may affect its section 79 safe-harbour position, but the precise legal consequence depends on the Act, Rules and facts. It is inaccurate to say that every compliance failure automatically proves criminal liability for the platform. Thirdly, the Rules' technical duties are not equivalent to a guarantee that every deepfake can be detected. Detection tools can produce false positives and false negatives; labels can be stripped or fail to travel across platforms; and provenance mechanisms may be absent from older or externally generated files.

The Rules therefore narrow one regulatory gap but do not settle the criminal-law question. They improve platform-level prevention, labelling and response. They do not by themselves provide a complete offence definition, a consistent culpability threshold for every harmful deepfake, or a compensation mechanism that restores the victim's reputation and control over their likeness.

In Shreya Singhal v Union of India, the Supreme Court struck down section 66A of the IT Act as unconstitutional because its vague and overbroad restrictions impermissibly burdened freedom of expression. The Court also read down the operation of section 79 in the context of intermediary liability, linking "actual knowledge" for removal to a court order or notification by the appropriate government or its agency concerning unlawful acts within the relevant constitutional framework. The case is not a deepfake judgment, but it is a central warning against responding to online harms through vague criminal prohibitions. A new offence framed around "false", "offensive" or "misleading" content without precise harm and intent requirements risks repeating the problems identified in section 66A. [6]

[6] Shreya Singhal v Union of India [2015] 5 SCC 1, [2015] INSC 257.


Copyright © Oxford Law Review. (www.oxfordlr.com)

*Oxford Law Review*, Vol. 1, Issue 1, 2026 DOI: https://10.5281/zenodo.23540582

In November 2023, Delhi Police registered a case following the circulation of an AI-manipulated video that placed actor Rashmika Mandanna's face onto footage originally featuring another person. Contemporary reporting stated that the FIR invoked sections 465 and 469 of the Indian Penal Code then in force, alongside sections 66C and 66E of the IT Act. The incident demonstrates how authorities may combine general offences with IT Act provisions to investigate a deepfake. It does not establish that all those provisions necessarily apply to every face-swap case, nor does the registration of an FIR amount to a judicial determination of guilt. It is best treated as an enforcement illustration, not binding precedent on the interpretation of deepfake liability. Because the BNS replaced the IPC for offences committed after commencement, the older IPC provisions should not be presented as the governing law for a new offence committed today.[7]

The Supreme Court's 2024 decision in Just Rights for Children Alliance v S Harish considered the legal treatment of child sexual exploitation material under the IT Act and POCSO. It underscores that online sexual exploitation must be analyzed through the exact statutory language, legislative purpose and conduct alleged. Its relevance is methodological as much as substantive: courts and investigators should not rely on colloquial labels such as "child pornography" or "deepfake" instead of identifying the statutory definition, the act performed, and the required intention. The decision does not eliminate the need for a careful separate analysis of synthetic-only images.[8]

There is no need to manufacture a deepfake-specific Supreme Court precedent to make the argument persuasive. The relevant legal questions can be developed from established principles on free expression, intermediary liability, electronic evidence, privacy and offence elements. A sound paper must distinguish binding holdings from analogies and reported police action. It should also be explicit where Indian appellate courts have not yet conclusively resolved a synthetic-media question. That candour strengthens, rather than weakens, the argument.

A criminal prohibition has limited value if investigators cannot preserve evidence or prove who did what. Deepfake cases present at least four evidentiary challenges.

The first is authentication. A clip may be edited, recompressed, cropped or forwarded through multiple services. The file submitted to police may not be the original file created by the accused. A forensic conclusion that the clip is manipulated is not identical to proof of when, where or by whom it was made.

The second is attribution. The uploader may use a pseudonymous account, a compromised account, a VPN, a shared device or an overseas service. Subscriber details and IP logs may assist but rarely prove authorship alone. Investigators should seek a combination of device artefacts, account logs, original files, message history, payment records, access logs and witness evidence, subject to lawful process and preservation constraints.

[7] 'Delhi Police Take Cognisance of Rashmika Mandanna's Deepfake Video, Begin Probe' The Indian Express (11 November 2023)

[8] Just Rights for Children Alliance v S Harish [2024] INSC



Copyright © Oxford Law Review. (www.oxfordlr.com)

*Oxford Law Review*, Vol. 1, Issue 1, 2026 DOI: https://10.5281/zenodo.23540582

The third is chain of custody. Screenshots can be useful for immediate reporting but may not preserve metadata or context. Victims should, where safe, preserve the original message, URL, account identifier, date and time, surrounding conversation and downloaded file without repeatedly forwarding the material. Investigators should record how each item was obtained, calculate and preserve cryptographic hashes where appropriate, maintain forensic images, and document every transfer and examination. Victim protection is crucial: evidence preservation should not require the victim to circulate intimate material further or expose it to unnecessary personnel.

The fourth is admissibility. The Bharatiya Sakshya Adhiniyam 2023 (BSA) governs electronic and digital records. Sections 61–63 address electronic or digital records and admissibility, with section 63 prescribing conditions for computer outputs and a statutory certificate framework. The correct certificate and proof requirements depend on how the record is produced and the circumstances of the case. An expert opinion may assist under the BSA's provisions on expert evidence, but an AI detector's output should not be treated as infallible or as a substitute for explaining its methodology, error rates, provenance and limitations.[9]

A further, counterintuitive problem is the "liar's dividend": once realistic fabrication becomes common, a wrongdoer can claim that genuine incriminating evidence is fake. Deepfake technology therefore harms not only people falsely depicted in fabricated content; it can also reduce confidence in authentic recordings. The law must avoid both gullibility and blanket scepticism. Authentication should be a structured inquiry into provenance, corroboration and competing explanations, not an assumption that a video is true because it looks real or false because AI exists.

The present framework has a distinctive pattern. It is comparatively well equipped where the deepfake is a means of committing an established offence. Fraudulent impersonation may be prosecuted as cheating by personation; qualifying explicit material may fall under the IT Act; extortion may be prosecuted under applicable BNS provisions; and qualifying imputations may attract defamation law. The difficult cases are those where the synthetic depiction itself causes serious harm but no existing offence clearly captures all relevant conduct.

Consider three examples. A scammer uses a cloned voice of a company director to induce an employee to transfer funds. The focus is not the synthetic voice in isolation but deception, personation and dishonest inducement. A person fabricates an intimate video of an adult and circulates it to humiliate the target. Depending on the facts, provisions on sexually explicit material, defamation, intimidation or other laws may apply, but a wholly synthetic depiction may not fit a provision framed around actual private-area imagery. A political actor circulates a realistic fabricated speech by a candidate. Depending on content, intent, election context and consequences, existing law may or may not apply; a general claim of misinformation does not itself establish a criminal offence.

[9] Bharatiya Sakshya Adhiniyam 2023, ss 39 and 61–63.


Copyright © Oxford Law Review. (www.oxfordlr.com)

*Oxford Law Review*, Vol. 1, Issue 1, 2026 DOI: https://10.5281/zenodo.23540582

The hidden pattern is that "deepfake" combines at least two separate legal problems: a representational wrong (the false attribution of words or acts) and a downstream harm (fraud, sexual humiliation, reputational injury, public disorder or electoral deception). Existing law tends to regulate the downstream harm rather than the technology or representational wrong as such. That is often desirable: it avoids criminalizing benign synthetic creativity. But it leaves a possible gap when non-consensual realistic impersonation causes serious dignitary harm without satisfying another offence.

The second-order effect is that an overbroad new offence could harm the very people it is meant to protect. Journalists may need to publish manipulated footage to expose manipulation; artists and satirists use impersonation; researchers test detection systems; and citizens may share content while debunking it. A law that criminalizes "creating or sharing a deepfake" without exceptions and mens rea safeguards could chill legitimate speech, give powerful people a tool to suppress criticism, or incentivise platforms to remove lawful material pre-emptively. The constitutional lesson from Shreya Singhal is that the definition of harm, fault and prohibited conduct must be precise.

A third pattern is the difference between removal and remedy. Rapid takedown reduces exposure, but it cannot undo downloads, private messages, reputational consequences or copies held elsewhere. Criminal prosecution may punish an offender but can be slow and may not deliver prompt restoration of the victim's agency. A coherent response therefore needs parallel tracks: evidence preservation and investigation; urgent and reviewable content restriction; victim support and civil remedies where available; and criminal prosecution when the elements are satisfied.

Three frameworks help organize reform without importing foreign law uncritically.
The first is harm-based criminalization. Criminal law should target demonstrable wrongs and culpable conduct, not a technology label. The more severe and foreseeable the harm, and the stronger the accused's intention or knowledge, the stronger the justification for criminal sanction. This framework supports targeted treatment of non-consensual sexual fabrication, fraud and deliberate harmful impersonation, while resisting blanket criminalization of synthetic media.

The second is proportionality and free-expression protection. Article 19(1)(a) protects expression, subject to the constitutionally permitted restrictions in Article 19(2). A new offence should be clear, necessary and proportionate to a recognized harm. It should distinguish deception intended to be believed from parody reasonably understood as parody, and it should avoid imposing strict criminal liability on a person who unknowingly forwards a file. A "public figure" should not lose all protection against fabricated sexual content, but public-interest criticism and satire require careful safeguards.[10]

The third is distributed responsibility. Deepfake creation and circulation may involve a model provider, application developer, user, platform, re-uploader and advertiser. Their

[10] Constitution of India 1950, art 19(1)(a), 19(2).


Copyright © Oxford Law Review. (www.oxfordlr.com)

*Oxford Law Review*, Vol. 1, Issue 1, 2026 DOI: https://10.5281/zenodo.23540582

knowledge, control and contribution differ. The creator may know the content is false; a platform may only learn of it after a complaint; a downstream user may be deceived by it. Responsibility should follow conduct and fault, not merely proximity to the technology. The 2026 IT Rules rightly place certain preventive duties on intermediaries, but those duties should be evaluated against feasibility, scale, privacy and error rates. A provider should not be treated as criminally liable simply because a user misused a general-purpose tool; conversely, deliberate facilitation after credible notice should not be insulated by formal labels.

International comparisons can be useful as design references, particularly on transparency, provenance and non-consensual intimate imagery. But a comparative paper must examine actual statutory text and implementation, not invoke "the EU approach" or "US law" as a single rule. India's constitutional structure, criminal procedure, linguistic diversity, platform ecology and enforcement capacity affect what will work. For this study, the strongest comparison is functional: what harm is addressed, what mental state is required, who bears duties, what remedy exists, and what safeguards prevent abuse?

Reform should proceed in a sequence rather than begin with a sweeping new offence. First, publish a consolidated prosecution and investigation protocol. It should map common fact patterns to potentially applicable provisions, explain the elements that must be proved, distinguish the IT Act from BNS offences, and caution against automatically charging identity theft or forgery whenever a face or voice is manipulated. Prosecutors should state the alleged offence and its statutory elements clearly. Such guidance improves consistency without changing the law.

Secondly, strengthen evidence preservation and forensic capacity. Police units need protocols for original-file preservation, lawful platform requests, metadata and log retention, hash recording, chain of custody, and expert examination. Training should include synthetic audio, image and video; but detection results must be expressed probabilistically and corroborated. Standardized reporting templates can help courts understand what an expert tested and what the test cannot establish.

Thirdly, implement the 2026 Rules with transparency and procedural safeguards. Platforms should provide accessible reporting channels, prioritise non-consensual intimate imagery and child exploitation, preserve evidence before removal where lawful and safe, and give reasons and appeal routes when content is removed. Automated detection should be audited for false positives, language and demographic performance, and evasion. Label and provenance requirements should not be represented as a guarantee of authenticity: a label may be absent because content predates the rule or passed through a system that strips metadata.

Fourthly, consider a narrowly defined statutory offence only after a gap analysis. A possible offence could target a person who, without consent and with intent to deceive, cause serious harm, or recklessly disregard a substantial and unjustifiable risk of such harm, creates or disseminates materially realistic synthetic audio-visual content falsely representing an identifiable person as engaging in specified conduct or making a statement. The provision should define material deception and serious harm, avoid criminalizing mere fictional depiction, and


Copyright © Oxford Law Review. (www.oxfordlr.com)

*Oxford Law Review*, Vol. 1, Issue 1, 2026 DOI: https://10.5281/zenodo.23540582

include carefully drafted defences or exclusions for consent, good-faith journalism, research, education, satire, parody and public-interest communication where the presentation is not intended to deceive or cause the prohibited harm. The exact mental-state threshold should be debated openly: requiring intent to cause serious harm may be underinclusive, while criminalizing mere negligence may be overbroad.

Fifthly, ensure the offence is not a shortcut around defamation or election law. If the purpose is merely to punish reputationally damaging falsehoods, existing defamation law and constitutional safeguards should remain relevant. A new offence should address the distinctive act of materially deceptive synthetic impersonation and require a defined level of culpability and harm. It should not empower officials to declare contested political speech criminal simply because it is inaccurate.

Sixthly, improve victim-centred remedies. Victims need prompt acknowledgement, preservation of evidence, a named point of contact, confidential handling of intimate material, protection from repeat circulation, and access to legal aid where appropriate. Any fast takedown process should permit urgent action while retaining a record for lawful investigation and providing a review mechanism to reduce malicious complaints. Removal, prosecution and compensation are distinct tools and should not be confused.

Finally, clarify synthetic child sexual abuse material through precise legislation if review reveals genuine ambiguity. The objective should be to protect children and prevent exploitation while ensuring that criminal provisions remain sufficiently clear to satisfy legality. Courts should not be left to resolve every new technical scenario through strained analogy.

India's response to AI-generated deepfakes should neither deny the seriousness of the harm nor exaggerate the existence of a legal vacuum. The IT Act 2000, BNS 2023, POCSO and related laws already cover important categories of conduct, including cheating by personation, certain uses of false electronic records, qualifying obscene or sexually explicit content, child exploitation, intimidation and defamation. The February 2026 amendments to the IT Rules add a significant layer of synthetic-media governance through prevention, labelling, provenance and removal duties.

The remaining challenge is coherence. A synthetic depiction may cause grave harm yet fall awkwardly between offences drafted for other conduct. The same technological form can be benign, expressive, deceptive or abusive; therefore, "deepfake" alone is not a sufficient basis for criminal liability. The right question is whether the prosecution can prove a defined act, the required mental state, a legally recognized harm and a reliable evidentiary connection to the accused.

The best path is a layered one: accurately apply existing offences; implement platform duties with safeguards; invest in forensic and investigative capacity; provide fast and reviewable victim remedies; and consider a narrow offence only if a demonstrated gap remains. Any new offence must be precise enough to survive constitutional scrutiny and restrained enough to preserve legitimate satire, journalism, art and research. In short, the problem is not simply that the law has failed to catch up with AI. It is that the law must distinguish carefully between synthetic


Copyright © Oxford Law Review. (www.oxfordlr.com)

*Oxford Law Review*, Vol. 1, Issue 1, 2026 DOI: https://10.5281/zenodo.23540582

representation and culpable deception, and between technological novelty and legally relevant harm.

## References

Information Technology Act 2000

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

Bharatiya Sakshya Adhiniyam 2023 (Act No. 47 of 2023)

Ministry of Information and Broadcasting, IT Rules 2021 (updated 10 February 2026)

Protection of Children from Sexual Offences Act 2012 (Act No. 32 of 2012)

Press Information Bureau, 'Government Strengthens Regulatory Framework to Address AI-Generated Deepfakes' (6 August 2026) <https://www.pib.gov.in/PressReleaseDetail.aspx?PRID=2295500&reg=6&lang=1>

The Indian Express, 'Delhi Police take cognizance of Rashmika Mandanna's deepfake video, begin probe' (11 November 2023) <https://indianexpress.com/article/cities/delhi/delhi-police-cognisance-rashmika-mandannas-deepfake-video-probe-9022585/>

*Shreya Singhal v Union of India* [2015] 5 SCC 1, [2015] INSC 257.

*Just Rights for Children Alliance v S Harish* [2024] INSC

## Publisher's Note

Oxford Law Review maintains neutrality with respect to the opinions, interpretations, and conclusions expressed in published articles, as well as jurisdictional claims in published maps and institutional affiliations. The views and arguments presented in this article are solely those of the author(s) and do not necessarily reflect the views of Oxford Law Review, its editors, or its publisher. Responsibility for the accuracy, originality, and integrity of the content rests with the author(s). Publication of an article does not constitute an endorsement by Oxford Law Review of the views expressed therein. All rights relating to the article are governed by the applicable publishing agreement between the author(s) or other rightsholder(s) and the publisher, together with applicable law.



Copyright © Oxford Law Review. (www.oxfordlr.com)