DEEPFAKE CRIMES AND INDIAN CRIMINAL LAW: ARE THE EXISTING LAWS SUFFICIENT?

by

Dr. Prashant Yadav,
Advocate

ABSTRACT

Deepfake crimes, spanning synthetic pornography, electoral disinformation, and financial endorsement fraud, pose an unprecedented threat to individual dignity and public order. Yet Indian criminal law lacks a dedicated statutory framework to address them. Instead, the state has managed the crisis through a patchwork of legally weightless executive advisories from the Ministry of Electronics and Information Technology (*MeitY*), generalised penal provisions in the Bharatiya Nyaya Sanhita ('BNS') and the Information Technology Act, 2000 ('IT Act') that were not designed for algorithmic generation, and synthetic-media amendments to the IT Rules that mandate disclosure rather than penalise harm. Consequently, victims have been forced to rely on expensive civil injunctions grounded in personality rights and privacy, exposing the inadequacy of the criminal justice system to protect ordinary citizens. This article audits the existing legal architecture to answer a pressing question: are current Indian laws sufficient to combat deepfake crimes? It concludes that they are fundamentally inadequate. It maps the collapse of intermediary safe harbour for generative outputs, the limits of disclosure-based IT Rules, and the public-law failure of governing by advisory, and it contributes the blueprint for a targeted criminal and regulatory legislative core.

Journal of Multi-Disciplinary Legal Research | Licensed under CC BY 4.0 — https://creativecommons.org/licenses/by/4.0/ | DOI: 10.5281/zenodo.21923016 | Title: Deepfake Crimes And Indian Criminal Law: Are The Existing Laws Sufficient?

1. Introduction

1.1 The Governance Gap: Advisories, Not Law

Late in 2023, a composited video appeared to show a well-known public figure endorsing a product she had never used, and a synthetic pornographic image of another celebrity, generated without her knowledge or consent, circulated widely enough to draw a rebuke from the Prime Minister.The scale of this threat has grown rapidly since these episodes: verification-industry data recorded a 280 per cent year-on-year increase in detected deepfake incidents in India in the first quarter of 2024 alone, and a subsequent industry report projected that deepfake-related fraud losses in India could reach approximately Rs 70,000 crore in 2024, with recorded deepfake cases having risen by 550 per cent since 2019[1]. Neither event was met with a criminal statute. It was met with a memorandum. Between December 2023 and March 2024, MeitY issued a rolling sequence of advisories to intermediaries under the Information Technology Act, 2000, rather than placing legislation before Parliament[2]. The pattern was unstable even on its own terms. The advisory of 1 March 2024 purported to require platforms to obtain the government's explicit prior permission before deploying an under-tested or unreliable artificial intelligence model to Indian users, a requirement that drew sharp criticism from start-up founders and industry counsel[3]. Within a

[1]Sumsub, 'Deepfake Cases Surge in Countries Holding 2024 Elections, Sumsub Research Shows' (Sumsub Newsroom, 30 May 2024) <https://sumsub.com/newsroom/deepfake-cases-surge-in-countries-holding-2024-elections-sumsub-research-shows/> accessed 13 August 2026; 'Deepfake Cases In India Surge 550% Since 2019, Projected Losses To Hit ₹70,000 Cr In 2024: Report' BW BUSINESSWORLD (5 December 2024) <https://www.businessworld.in/article/indias-deepfake-cases-up-550-losses-may-hit-rs-70000-cr-by-2024-report-541202> accessed 13 August 2026.

[2]See Advisory, Ministry of Elecs. & Info. Tech., Due Diligence by Intermediaries/Platforms under the Information Technology Act, 2000 and Rules thereunder (Mar. 1, 2024) (India); Advisory, Ministry of Elecs. & Info. Tech., Due Diligence by Intermediaries/Platforms under the Information Technology Act, 2000 and Rules thereunder (Mar. 15, 2024) (India) (superseding the March 1 advisory); see also New Advisory of MeitY: AI Platforms Don't Need Government Permission, Focus on Deepfakes, BUS. TODAY (Mar. 16, 2024), https://www.businesstoday.in/tech-today/news/story/new-advisory-of-meity-ai-platforms-dont-need-government-permission-focus-on-deepfakes-421703-2024-03-16.

[3]S&R Law (SNR), Investing in AI in India (Part 3): AI-Related Advisories Under the Intermediary Guidelines 2–3 (Oct. 2024), https://www.snrlaw.in/wp-content/uploads/2024/10/SR-AI-Focus-Investing-in-AI-in-India-Part-3-AI-related-Advisories-Under-the-Intermediary-Guidelines.pdf.

fortnight MeitY withdrew that requirement, replacing it on 15 March 2024 with a revised advisory that abandoned prior government approval altogether and instead directed intermediaries to label content susceptible to deepfake misuse and to embed it with traceable metadata identifying the platform or computer resource used to create or modify it[4]. Both instruments were issued as advisories addressed to platforms under the Intermediary Guidelines, not as delegated legislation notified in the Gazette, and neither created a penal offence. This sequence, permission demanded then withdrawn, obligation restated then softened, establishes the pattern this article calls *governing by advisory*: the substitution of a revisable and legally unenforceable executive instrument for the deliberative process of statute, precisely at the moment when synthetic media was beginning to cause dignitary and financial harm to identifiable people.

1.2 The Argument: Four Audits Converging on Insufficiency

This article's thesis is direct: India's existing legal framework is fundamentally insufficient to address deepfake crimes because it relies on a legally exhausted strategy of executive advisories, ill-fitting civil remedies, and an outdated intermediary liability regime[5]. The article develops this thesis through four sequential doctrinal audits. Section 2 demonstrates the statutory void: the BNS and the IT Act supply penal provisions built for a human impersonator or a human publisher, not for a generative model that composes a synthetic likeness the accused never physically manipulated, and the state has attempted to fill that gap with advisories that delegated-legislation doctrine will not sustain. Section 3 exposes the civil-law band-aid: the High Courts have protected celebrities against synthetic exploitation through personality-rights injunctions, but that remedy is

[4]New Advisory of MeitY, supra note 1.

[5]Cf. Legal Blur, Deepfakes and Indian Law: Copyright and Liability (July 11, 2026), https://legalblur.com/deepfakes-and-indian-law-copyright-liability/ (arguing for a codified personality-rights statute, a tiered intermediary liability model, and a standalone criminal offence in the Bharatiya Nyaya Sanhita).

structurally unavailable to the ordinary victim of non-consensual synthetic pornography. Section 4 traces the collapse of Section 79 safe harbour when a generative platform composes, rather than hosts, the offending content. Section 5 examines the limits of the 2026 synthetic-media amendments to the IT Rules, which regulate an artefact's provenance rather than the harm its circulation causes. Sections 6 and 7 draw on comparative practice, the criminal statutes of Virginia and Texas, the federal TAKE IT DOWN Act, and the EU AI Act's transparency model, to propose the targeted legislative core the four audits show is missing.

2. The Statutory Void: Why Existing Offences Do Not Reach Algorithmic Generation

2.1 Cheating, Personation and Obscenity: Provisions Not Built for AI

The provisions currently pressed into service against deepfakes were drafted for a different actor. Section 319 of the BNS punishes cheating by personation, pretending to be another person, substituting one person for another, or representing that a person is someone other than they really are, with imprisonment of up to five years[6]. Alongside it sit Sections 75, 77, 351, 353 and 356 of the BNS, addressing sexual harassment, voyeurism, criminal intimidation, statements conducing to public mischief, and defamation, together with Sections 66C, 67 and 67A of the IT Act, which punish identity theft and the publication or transmission of obscene or sexually explicit material in electronic form[7]. Section 66C punishes fraudulent or dishonest use of another person's electronic signature, password, or unique identification feature with imprisonment of up to three

[6]Bharatiya Nyaya Sanhita, 2023, No. 45 of 2023, § 319 (India). Section 319 restates, with an enhanced maximum sentence, the offence formerly found at section 416 of the Indian Penal Code, 1860. See Section 319 BNS (Bharatiya Nyaya Sanhita) 2023: Cheating by Personation, TESTBOOK, https://testbook.com/judiciary-notes/section-319-bns (last visited Aug. 13, 2026).

[7]Bharatiya Nyaya Sanhita, 2023, No. 45 of 2023, §§ 75, 77, 351, 353, 356 (India); Information Technology Act, 2000, No. 21 of 2000, §§ 66C, 67, 67A (India); see also Key Provisions Related to Cyber Crimes Under Bharatiya Nyaya Sanhita, 2023 (BNS), LAWSECTION.IN (Mar. 31, 2026), https://lawsection.in/cyber-crimes-under-bharatiya-nyaya-sanhita-2023-bns/.

years and a fine of up to one lakh rupees; Section 67 punishes publication of obscene material in electronic form with imprisonment of up to three years on first conviction; Section 67A extends the maximum to five years where the material is sexually explicit[8]. Each provision, however, was drafted for a human impersonator typing a fraudulent message, a human uploader transmitting an obscene photograph, or a human speaker uttering a defamatory statement. None was drafted for a generative model that composes, from a text prompt and a training corpus, a synthetic likeness the accused never physically manipulated, filmed, or photographed. The statutory language of "publishing," "transmitting," and "personation" assumes a single human actor whose conduct and content coincide; a diffusion model interposes an architecture and a corpus between the prompter's conduct and the artefact's content, and none of Sections 319, 351, 353, 356, 66C, 67 or 67A was written with that interposition in view.

2.2 The Mens Rea Problem: Assigning Criminal Intent to Algorithmic Design

The doctrinal failure at the centre of this audit is narrower and more precise than a general observation that the law is outdated: traditional criminal defamation, obscenity, and identity-theft standards require a human mental state directed at a specific victim, but a generative model's output is the product of a prompt, a training corpus, and a model architecture, none of which maps cleanly onto either the person who typed the prompt or the entity that built the model. A prosecutor seeking to establish, under Section 356 of the BNS, that an accused "intend[ed] to harm, or kn[ew] or ha[d] reason to believe that such imputation will harm, the reputation" of a person[9], or, under Section 67A of the IT Act, that an accused knowingly published sexually explicit material, must locate that intent in a specific human act. Where the prompter supplies only a general instruction

[8]Information Technology Act, 2000, No. 21 of 2000, §§ 66C, 67, 67A (India).
[9]Bharatiya Nyaya Sanhita, 2023, No. 45 of 2023, § 356 (India).

and the model's own weights determine the specific likeness, pose, or statement generated, the causal chain between the prompter's charged mental state and the artefact's specific content is attenuated in a way the drafters of Sections 356 and 67A did not contemplate.

This article maintains a distinction the underlying research brief for this audit rightly demands: between the prompter's criminal intent and the developer's product-design responsibility. Collapsing the two would produce either of two undesirable results. Treating the model's autonomy as a defence would let a prompter who deliberately engineered a harmful output escape liability by pointing to the stochastic layer between instruction and artefact. Conversely, treating the developer as strictly liable for every downstream misuse would expose a model's designer to criminal liability for uses it neither directed nor could reasonably have foreseen, a result at odds with the fault-based structure of the BNS itself. No dedicated Indian secondary source disaggregating prompter mens rea from developer design liability in doctrinal terms was located over the course of the research for this article. The analysis that follows is accordingly developed from first principles of BNS mens rea doctrine, specifically the requirement, common to Sections 319, 353 and 356, that the accused act with knowledge or intent directed at a describable consequence, cross-checked against the design-stage, foreseeability-based standard of care used in product liability analysis. On that synthesis, a prompter who engineers, iterates, or selects among outputs to produce a specific deceptive or harmful likeness satisfies the knowledge or intent element the BNS already requires; a developer's liability, by contrast, is better addressed prospectively, as Section 4.2 of this article proposes, through a reasonable-care standard fixed at the point of model design rather than retrofitted onto offences that presuppose a single culpable human author.

2.3 Governing by Advisory: The MeitY Notifications and Their Legal Weightlessness

The public-law objection to the pattern described in Section 1.1 is not merely that advisories are informal; it is that they cannot lawfully do what MeitY has used them to do. An advisory issued under the Intermediary Guidelines cannot create a penal offence, because delegated legislation, and a fortiori a non-statutory advisory that is not even notified as delegated legislation, must operate within the four corners of its parent Act. The Supreme Court held as much in *Indian Express Newspapers (Bombay) Pvt. Ltd. v. Union of India*, ruling that subordinate legislation which exceeds or conflicts with its enabling statute may be struck down as manifestly arbitrary or ultra vires[10], a principle reaffirmed two decades later in *Confederation of Ex-Servicemen Associations v. Union of India*[11]. Digital-rights commentary has pressed precisely this objection against MeitY's advisory practice. The rule-making power conferred by Section 87(1) of the IT Act is confined to framing rules that "carry out the provisions" of the Act; Section 87(2)(zg) authorises rules for intermediary guidelines under Section 79(2), and Section 87(2)(z) authorises rules for blocking procedures under Section 69A(2)[12]. Neither provision authorises MeitY to issue free-standing advisories, clarifications, or directions that intermediaries must obey as a condition of retaining safe harbour. That objection has since acquired sharper teeth. On 30 March 2026, MeitY published draft amendments to the IT Rules that would insert a new Rule 3(4), making intermediary compliance with MeitY-issued "clarifications, advisories, directions, SOPs, codes of practice, and guidelines" a condition of safe harbour under Section 79, without anchoring that obligation to any rule-making power the IT Act actually confers[13]. Commentary on the draft has

[10]Indian Express Newspapers (Bombay) Pvt. Ltd. v. Union of India, (1985) 1 S.C.C. 641 (India).

[11]Confederation of Ex-Servicemen Ass'ns v. Union of India, (2006) 8 S.C.C. 399 (India).

[12]Information Technology Act, 2000, No. 21 of 2000, § 87(1), (2)(z), (2)(zg) (India); see Internet Freedom Foundation, Sound the Alarm: IFF's First Read on MeitY's Draft IT Rules Second Amendment, 2026 (Mar. 30, 2026), https://internetfreedom.in/sound-the-alarm-iffs-first-read-on-meitys-draft-it-rules-second-amendment-2026/.

[13]Internet Freedom Foundation, supra note 8 (describing the Draft Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Second Amendment Rules, 2026, published for comment on 30 March 2026, with the comment period closing 14 April 2026).

argued that this would convert exactly the species of legally weightless instrument this article describes into a binding precondition of a statutory immunity, an outcome *Indian Express Newspapers* and *Confederation of Ex-Servicemen Associations* suggest the rule-making power under Section 87 cannot support[14].Commentary on the draft has also observed that Rule 3(4) may not create a wholly new executive power so much as formalise a compliance dynamic that already existed in practice, since intermediaries facing the cost and uncertainty of litigation have every incentive to treat an advisory as binding even where it lacks that character in law; that MeitY was required to extend its initial comment window after industry and civil-society pushback is itself indicative of how contested the underlying claim to rule-making authority has proved[15]. The Supreme Court's reading-down of Section 79(3)(b) in *Shreya Singhal v. Union of India*, discussed further in Section 4.1 below, already requires that "actual knowledge" of unlawful content come through a court order or a government notification issued under law[16]; an advisory that is neither adds a further, unauthorised layer of executive discretion onto a threshold the Court has already fixed. Governing by advisory, in short, is not simply an underuse of the legislative process. It is, on the reasoning of India's own administrative law, a form of governance the Constitution's allocation of law-making power does not permit.

3. The Civil-Law Band-Aid: Personality Rights as a Privatised Remedy

[14]See id.; Guest Post: Digital Structural Authoritarianism, India's IT Rules 2021 Draft Amendment, INDIAN CONST. L. & PHIL. BLOG (Apr. 6, 2026), https://indconlawphil.wordpress.com/2026/04/06/guest-post-digital-structural-authoritarianism-indias-it-rules-2021-draft-amendment/comment-page-1/.

[15]'Advisories, Safe Harbour And Executive Power: Rethinking Draft 2026 IT Rules' LIVELAW (7 July 2026) <https://www.livelaw.in/lawschool/articles/advisories-safe-harbour-executive-power-rethinking-draft2026-it-rules-539958> accessed 13 August 2026; see also 'IT Rules Second Amendment 2026: DPDP Collision Explained' CANDOUR LEGAL (28 April 2026) <https://candourlegal.com/it-rules-second-amendment-2026-dpdp-collision/> accessed 13 August 2026 (noting the consultation period was extended from the original 14 April 2026 deadline following industry and civil-society representations).

[16]Shreya Singhal v. Union of India, (2015) 5 S.C.C. 1 (India).

3.1 Anil Kapoor v. Simply Life India: Injunctive Relief for the Famous

In the absence of a criminal remedy, India's High Courts have improvised a civil one. On 20 September 2023, the Delhi High Court granted actor Anil Kapoor a sweeping ex parte omnibus injunction protecting his name, image, voice, and persona against unauthorised commercial exploitation, including deepfakes generated using artificial intelligence that depicted him as other film characters and morphed images placing his face onto pornographic content[17]. Justice Prathiba M. Singh's order rested on three distinct legal bases: personality rights as an extension of privacy and dignity under Article 21 of the Constitution, copyright in Kapoor's dialogue and associated cinematic works, and the common-law tort of passing off[18]. None of these three bases is a criminal-law remedy. The order directed domain registrars to lock infringing domain names and directed MeitY and the Department of Telecommunications to block the identified URLs, but it did not, and under the causes of action pleaded could not, expose any defendant to prosecution, a fine payable to the state, or a criminal record. It became, nonetheless, the template Indian courts have since followed whenever a well-known plaintiff's likeness is misappropriated through generative tools.

3.2 Arijit Singh v. Codible Ventures: Extending the Template to Voice

The template's extension to synthetic audio confirms both its reach and its limits. On 26 July 2024, the Bombay High Court granted playback singer Arijit Singh an ex parte ad interim injunction restraining a group of AI voice-cloning platforms from using his name, voice, vocal style, and

[17] Anil Kapoor v. Simply Life India, CS(COMM) 652/2023 & I.A. 18237–18243/2023 (Del. H.C. Sept. 20, 2023) (India); see Software Freedom Law Ctr., India, Anil Kapoor v. Simply Life India, https://sflc.in/litigation/anil-kapoor-v-simply-life-india/ (last visited Aug. 13, 2026).

[18] Anil Kapoor v. Simply Life India, CS(COMM) 652/2023, at 6–12 (Del. H.C. Sept. 20, 2023) (India); see also Anil Kapoor v. Simply Life India & Ors. (2023): Personality Rights in the Age of Artificial Intelligence, RECORD OF LAW (Mar. 28, 2026), https://recordoflaw.in/anil-kapoor-v-simply-life-india-ors-2023-personality-rights-in-the-age-of-artificial-intelligence/.

likeness to generate synthetic recordings[19]. Justice R.I. Chagla observed that AI tools capable of replicating a performer's voice and persona leave considerable room for misutilisation by unscrupulous actors, and found that the plaintiff had made out a prima facie case for interim relief[20]. As in *Anil Kapoor*, however, the relief issued through a civil suit invoking moral rights under Section 38-B of the Copyright Act, 1957 and common-law personality rights, not any provision of the BNS or the IT Act[21]. That two High Courts, in successive years, reached for the same civil toolkit when confronted with materially different synthetic harms, a morphed pornographic image in one case, a cloned singing voice in the other, illustrates how readily personality-rights doctrine has absorbed the deepfake problem in the absence of any criminal alternative, and how consistently the resulting relief has remained confined to injunctions running against named or joined defendants rather than public prosecutions.

The template identified in this Section has continued to extend to further public figures since these two decisions. Reporting on India's deepfake landscape has noted that performers including Aishwarya Rai Bachchan and the actor Jr NTR have likewise obtained court orders restraining the unauthorised synthetic use of their name, image, and likeness, confirming that the personality-rights route has become the default recourse available to any well-known Indian figure confronted

[19] Arijit Singh v. Codible Ventures LLP, I.A.(L) No. 23560 of 2024 in Com. IPR Suit (L) No. 23443 of 2024 (Bom. H.C. July 26, 2024) (India).

[20] See Using His Name, Voice & Likeliness in Form of AI Content Leaves Room for Misutilization of Such Tools: Bombay HC Protects 'Personality Rights' of Arijit Singh, VERDICTUM (Feb. 20, 2025), https://www.verdictum.in/court-updates/high-courts/bombay-arijit-singh-v-codible-ventures-llp-personality-rights-artificial-intelligence-1546255; AZB & Partners, Bombay High Court Safeguards Celebrity Personality Rights in AI Era: Arijit Singh Case (2024), https://www.azbpartners.com/bank/bombay-high-court-safeguards-celebrity-personality-rights-in-ai-era-arijit-singh-case/.

[21] See Bombay HC Grants Ad-interim in Favour of Singer Arijit Singh Against Unauthorised Commercial Exploitation, LAWFUL TALKS, https://www.lawfultalks.net/news/bombay-hc-grants-ad-interim-in-favour-of-singer-arijit-singh-against-unauthorised-commercial-exploitation (last visited Aug. 13, 2026).

with synthetic exploitation, rather than an exceptional remedy confined to the two decisions discussed above[22].

3.3 Why the Remedy Fails Ordinary Victims of Non-Consensual Synthetic Pornography

The distributive weakness of this civil template is structural, not incidental. Personality-rights litigation of the *Anil Kapoor* and *Arijit Singh* kind is available only to plaintiffs who can plead three things at once: a recognisable commercial persona whose goodwill the court can protect, the resources to retain counsel capable of drafting and moving an ex parte application against dozens of joined defendants, and a court willing to treat the plaintiff's name and likeness as an asset with established commercial value[23]. None of those three conditions describes the ordinary person victimised by non-consensual synthetic pornography. Such a victim typically has no commercial persona to protect, since the harm is dignitary rather than economic; the fee structures that make ex parte commercial-suit litigation viable for a film star are simply unavailable to a private individual; and a court asked to protect an ordinary person's likeness has no established line of authority, built as it is on celebrity plaintiffs' publicity interests, to draw on.The scale of this gap is considerable and markedly gendered: one 2026 industry study found that ninety-three per cent of deepfake victims globally are women, that cybercrime complaints in India involving women rose from approximately 50,000 in 2024 to nearly 80,000 by 2026, and that sixty-two per cent of deepfake abuse cases involving women go unreported, largely on account of stigma rather than an absence of harm. Given that ordinary, non-celebrity women make up the overwhelming majority of this underreported and unremedied population, the structural inaccessibility of the personality-

[22]'India's Deepfake Crisis: Women Are Falling Prey to AI Menace More Than Men' DIGIT.IN (10 March 2026) <https://www.digit.in/features/general/indias-deepfake-crisis-women-are-falling-prey-to-ai-menace-more-than-men.html> accessed 13 August 2026.

[23]See Anil Kapoor v. Simply Life India & Ors. (2023): Personality Rights in the Age of Artificial Intelligence, supra note 15.

rights route is not a marginal limitation on an otherwise adequate remedy but the central reason it fails the class of victim most affected by non-consensual synthetic pornography[24]. The civil template that has worked twice for Bollywood and twice for the recording industry therefore offers this most common and most severe class of deepfake victim no comparable route to relief at all. It is a band-aid sized for the famous, applied to a wound that, statistically, is overwhelmingly suffered by the anonymous.

4. The Section 79 Problem: When the Platform Is the Author, Not the Host

4.1 Safe Harbour's Premise and Why Generative Composition Breaks It

Section 79 of the IT Act exempts an intermediary from liability for third-party information that it merely transmits, hosts, or stores, conditioned on due diligence and on expeditious action once it receives actual knowledge of unlawful content[25]. In *Shreya Singhal v. Union of India*, the Supreme Court read down Section 79(3)(b) to require that such "actual knowledge" arrive through a court order or a notification from the appropriate government, precisely so that intermediaries would not be forced into private adjudication of every complaint of unlawfulness[26]. That structure presupposes a passive intermediary: an entity that stores or carries content authored by someone else. A generative AI system offering a deepfake-creation tool does not fit that premise. It does not host a third party's pre-existing file; it composes new synthetic media in response to a prompt, using a model whose parameters, and therefore whose expressive choices, the platform itself

[24]'India's Deepfake Crisis: Women Are Falling Prey to AI Menace More Than Men' DIGIT.IN (10 March 2026) <https://www.digit.in/features/general/indias-deepfake-crisis-women-are-falling-prey-to-ai-menace-more-than-men.html> accessed 13 August 2026; see also '62% of Deepfake Abuse Cases Involving Women Go Unreported Due to Stigma: pi-labs Report' MEDIANEWS4U (10 March 2026) <https://www.medianews4u.com/62-of-deepfake-abuse-cases-involving-women-go-unreported-due-to-stigma-pi-labs-report/> accessed 13 August 2026.

[25]Information Technology Act, 2000, No. 21 of 2000, § 79 (India).

[26]Shreya Singhal v. Union of India, (2015) 5 S.C.C. 1 (India).

designed and trained. That places the service structurally closer to a publisher exercising editorial control over content it produces than to a conduit carrying content it did not create, a distinction the Government of India has itself now acknowledged turns on the specific function an AI service performs rather than on any blanket classification. Responding to a question in the Lok Sabha, the Minister of State for Electronics and Information Technology confirmed that whether a given AI system falls within the definition of "intermediary" under Section 2(1)(w) of the IT Act, and is therefore eligible for the Section 79 exemption, depends on the nature of the service and the functions the system performs[27]. On the government's own function-based test, a service whose function is the algorithmic composition of a deceptive likeness is not obviously hosting anything at all.

The dispute now before the Bombay High Court illustrates the stakes of that ambiguity in real time. Union Minister Nitin Gadkari filed a civil suit against Meta Platforms, X Corp, and Google, along with unidentified "Ashok Kumar" defendants, after AI-generated videos and manipulated images falsely portrayed him as personally profiting from the government's E20 ethanol-blending policy[28]. A bench led by Justice Abhay Ahuja permitted the suit to proceed on 27 July 2026, and on 5 August 2026 Justice Arif Doctor directed the platforms to remove the identified deepfake content immediately and to disclose the subscriber data of the anonymous accounts responsible

[27]Generative AI Safe Harbour Status Dependent on Function: Government, OUTLOOK INDIA (July 2026), https://www.outlookindia.com/national/generative-ai-safe-harbour-status-dependent-on-function-government; see also Govt Says Eligibility of Generative AI Services for Safe Harbour Depends on Nature of Service, INDIA GAZETTE, https://www.indiagazette.com/news/279215069/govt-says-eligibility-of-generative-ai-services-for-safe-harbour-depends-on-nature-of-service (last visited Aug. 13, 2026).

[28]Nitin Jairam Gadkari v. Meta Platforms, Inc., I.A.(L) No. 25165 of 2026 (Bom. H.C. filed July 27, 2026) (India); see Can Gadkari Hold Meta and Google Liable for Deepfakes? Case Revives Safe Harbour Questions, OUTLOOK BUS. (July 2026), https://www.outlookbusiness.com/deeptech/can-gadkari-hold-meta-and-google-liable-for-deepfakes-case-revives-safe-harbour-questions.

for uploading it[29]. The order is a defamation-and-disclosure remedy, not an adjudication that the platforms' own generative tools composed the offending content, and it therefore does not resolve the doctrinal question this article raises. But it confirms that the safe-harbour question is no longer academic: a sitting Union Minister has found it necessary to litigate platform accountability for synthetic content through the same civil route this article's Section 3 shows is unavailable to an ordinary victim, underscoring the urgency of a criminal-law and regulatory answer that does not depend on the resources or standing to bring a civil suit against three multinational platforms at once.

4.2 Toward a Reasonable-Care Standard for Model Design

This article's structural prescription for Section 79 rejects the binary choice between strict liability and full immunity that has framed the debate so far. In its place it proposes a reasonable-care standard assessed at the point of model design and deployment: did the generative system incorporate readily available safeguards against producing non-consensual sexual imagery of an identifiable person, or against generating identifiable political disinformation, before it was made available to Indian users? This design-stage inquiry is distinct from, and prior to, a platform's post-hoc takedown obligations under the amended IT Rules discussed in Section 5. Commentary proposing a tiered removal and liability framework tied to the type of synthetic content at issue points in a compatible direction, recommending, for example, that content constituting non-consensual intimate imagery or child sexual abuse material be subject to the shortest removal windows and the sharpest liability consequences, with graduated treatment for other categories of

[29]Indian Court Forces Meta, Google to Reveal Who Made Deepfakes of Cabinet Minister, TECHTIMES (Aug. 6, 2026), https://www.techtimes.com/articles/323273/20260806/indian-court-forces-meta-google-reveal-who-made-deepfakes-cabinet-minister.htm; Bombay HC Directs X, Meta, Google to Take Down Defamatory AI Deepfakes on Gadkari, DECCAN CHRON. (Aug. 2026), https://www.deccanchronicle.com/nation/bombay-hc-directs-x-meta-google-to-take-down-defamatory-ai-deepfakes-on-gadkari-1976879.

synthetic harm[30]. A reasonable-care standard borrows that graduated logic but relocates it upstream, to the design and training decisions a generative-model provider makes before a single output is ever generated, rather than leaving it entirely to what a platform does after a specific piece of synthetic content has already caused harm.

4.3 Engaging the Counter-Argument: Would Reasserting Liability Chill Innovation?

The strongest objection to reforming Section 79 along these lines is that treating a generative-model provider as a composer rather than a host would expose India's nascent AI development sector to precisely the liability risk safe harbour was designed to prevent for early internet platforms, discouraging investment and innovation in a strategically important industry. The 2026 amendments to the IT Rules, which recast intermediary due-diligence obligations for AI-generated content, have already drawn this criticism from industry commentary concerned about compressed compliance timelines and open technical questions around metadata persistence[31]. This article's answer is to distinguish, sharply, an ex ante design-stage duty of reasonable care from ex post liability for every downstream misuse of a deployed model. A design-stage standard targets a narrow and foreseeable category of high-severity harms, principally the generation of non-consensual sexual imagery of identifiable people and identifiable political disinformation, and asks only whether the provider incorporated safeguards against those specific, foreseeable outputs that were reasonably available to it at the time of deployment. It does not impose liability for every unforeseeable way a user might later misuse a general-purpose model, which is the feature of a strict-liability regime that legitimately concerns industry. A generative-AI developer that has taken

[30]Legal Blur, supra note 6 (proposing a three-tier intermediary liability model tied to content category, alongside a graduated Bharatiya Nyaya Sanhita offence).

[31]See AI Generated Content, Safe Harbour and Due Diligence: How the 2026 IT Rules Recast Intermediary Obligations in India, MONDAQ (2026), https://www.mondaq.com/india/new-technology/1760556/ai-generated-content-safe-harbour-and-due-diligence-how-the-2026-it-rules-recast-intermediary-obligations-in-india.

reasonable, industry-standard steps to prevent its model from producing non-consensual intimate imagery of a real, identifiable person has nothing to fear from this standard even where a determined user circumvents those safeguards; a developer that has taken no such steps, by contrast, is not the kind of innovator safe harbour was designed to protect.

5. The Synthetic-Media Amendments: Disclosure Is Not Deterrence

5.1 Labelling, Watermarking and the IT Rules' Provenance Model

The actual regulatory instrument now in force is the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, notified by MeitY on 10 February 2026 via Gazette Notification G.S.R. 120(E) and brought into effect from 20 February 2026, exercising the rule-making power conferred by Sections 87(1), 87(2)(z), and 87(2)(zg) of the IT Act[32]. The Amendment Rules define "synthetically generated information" as audio, visual, or audio-visual content that has been artificially created or algorithmically altered using a computer resource in a manner that makes it appear real, and require intermediaries to obtain a user declaration on whether uploaded content is synthetically generated, to deploy technical measures to verify that declaration where feasible, and, where content is confirmed synthetic, to ensure it is prominently labelled and embedded with permanent, non-removable metadata or a comparable provenance mechanism[33].In practical terms, this obligation has been fixed at a three-hour window running from the point a platform receives notice of unlawful synthetically generated content,

[32]Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, G.S.R. 120(E) (Feb. 10, 2026) (India); Government Notifies Information Technology Amendment Rules 2026, CHAMBERS & PARTNERS (Feb. 22, 2026), https://chambers.com/articles/government-notifies-information-technology-amendment-rules-2026.

[33]India's IT Amendment Rules 2026 Explained: How the New Law Regulates Deepfakes and AI-Generated Content, THE LEGAL HUBB (July 6, 2026), https://thelegalhubb.com/indias-it-amendment-rules-2026-explained-how-the-new-law-regulates-deepfakes-and-ai-generated-content/; Regulating Synthetically Generated Information: India's IT Rules Amendment of 2026, RNA TECH. & IP ATT'YS (Mar. 6, 2026), https://rnaip.com/regulating-synthetically-generated-information-indias-it-rules-amendment-of-2026/.

among the shortest compliance timelines imposed on intermediaries by any jurisdiction currently regulating synthetic media[34]. The Amendment Rules also compress takedown timelines for unlawful synthetic content into tightly tiered windows, and non-compliance risks forfeiture of the Section 79 safe harbour discussed in Section 4[35]. What the Amendment Rules do not do is create any offence. They tell the viewer what an artefact is; they say nothing about what happens to the person who made it, or the person it depicts, once the label has been affixed. A regime that requires a non-consensual synthetic pornographic image to be labelled "synthetically generated" before it is left online, rather than criminalising its creation, addresses the artefact's provenance while leaving its harm to the victim's reputation, dignity, and safety entirely unremedied.

5.2 The EU AI Act Comparator: Transparency Without Criminalisation

The European Union's Artificial Intelligence Act supplies the comparative baseline this audit needs, because it represents the most developed transparency regime for synthetic media currently in force anywhere. Article 50 requires providers of generative AI systems to ensure that outputs are marked in a machine-readable format as artificially generated, and separately requires deployers who use an AI system to produce a deepfake to disclose that the content has been artificially generated or manipulated[36]. Article 50(4) narrows that disclosure duty where the

[34]"India's IT Rules 2026 Focus on AI Content Takedowns and Oversight' OPEN (10 April 2026) <https://openthemagazine.com/india/indias-new-it-rules-2026-focus-on-ai-content-takedowns-and-oversight> accessed 13 August 2026 (noting the three-hour removal window applicable to Synthetically Generated Information from 20 February 2026); see also 'India Supreme Court Demands Criminal Law for Digital Arrest, Deepfake Fraud' TECHTIMES (29 July 2026) <https://www.techtimes.com/articles/321946/20260729/india-supreme-court-demands-criminal-law-digital-arrest-deepfake-fraud.htm> accessed 13 August 2026.

[35]See Managing Intell. Prop., Regulating Synthetically Generated Information: India Amends IT Intermediary Rules (2026), https://www.managingip.com/article/2g3bz6kdk7qxkoxxcsg00/sponsored-content/regulating-synthetically-generated-information-india-amends-it-intermediary-rules.

[36]Regulation 2024/1689, art. 50(2), (4), of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), 2024 O.J. (L 1689) (EU); see Article 50: Transparency Obligations for Providers and Deployers of Certain AI Systems, EU ARTIFICIAL INTELLIGENCE ACT, https://artificialintelligenceact.eu/article/50/ (last visited Aug. 13, 2026).

deepfake forms part of an evidently artistic, creative, satirical, or fictional work, limiting the obligation to disclosing the existence of the manipulated content in a manner that does not hamper the display or enjoyment of the work[37].The European regime backs this transparency mandate with substantial administrative penalties for non-compliance, empowering regulators to impose fines running up to thirty-five million euros or seven per cent of a provider's global annual turnover, whichever is higher; India's Amendment Rules attach no comparable monetary penalty to a failure to label synthetic content, relying instead on the indirect threat of safe-harbour forfeiture discussed in Section 4[38]. That free-expression carve-out is an important comparative point for the safeguards this article proposes in Section 7. But Article 50, for all its comparative sophistication, remains a transparency instrument embedded within a product-safety regulation. It does not create a criminal offence for producing a harmful deepfake; it requires that the deepfake be labelled. Even the European Union's more developed framework, in other words, leaves the criminalisation question to individual member states, which confirms that India's IT Rules are not an outlier in stopping at disclosure. They are, however, an outlier in stopping at disclosure without any accompanying member-state-style criminal backstop, since India, unlike an EU member state, has no separate national deepfake offence sitting alongside its transparency rule.

5.3 Brittle Against Circumvention: Offshore Tools and the Limits of Platform-Side Fixes

A structural weakness closes this audit of the IT Rules. Labelling and metadata obligations bind intermediaries operating in, or targeting, the Indian market; they do not bind a deepfake generation

[37]Regulation 2024/1689, art. 50(4), 2024 O.J. (L 1689) (EU); Transparency Obligations Under Article 50 of the AI Act, EUR. COMM'N: DIGITAL STRATEGY, https://digital-strategy.ec.europa.eu/en/faqs/transparency-obligations-under-article-50-ai-act (last visited Aug. 13, 2026).

[38]'Deepfake Statistics [2026]: Growth, Fraud & Detection Data' STATIONX <https://app.stationx.net/articles/deepfake-statistics> accessed 13 August 2026 (summarising the penalty structure attaching to the transparency obligations under the EU Artificial Intelligence Act).

tool with no Indian nexus at all. A synthetic pornographic image or a fabricated endorsement video generated using an offshore tool, then uploaded through a channel that strips embedded metadata on re-upload, a routine consequence of routine image compression and social-media re-encoding, defeats the provenance model entirely, without the uploader ever violating the Amendment Rules as such, because the Rules bind the Indian intermediary's labelling conduct, not the offshore generation step[39]. This is not a speculative or uniquely Indian risk. It mirrors the documented persistence of unlabelled AI-generated content during the 2024 European Parliament elections, notwithstanding the EU AI Act's own disclosure requirement, which demonstrates that a provenance-only regime, however well drafted, cannot by itself prevent circumvention once content crosses a jurisdictional boundary or passes through a re-upload pipeline that discards metadata.Independent monitoring by researchers tracking the 2024 European Parliament campaign identified 131 pieces of undeclared AI-generated or AI-manipulated content circulating across major platforms in the weeks before that vote, confirming that the labelling obligation was widely disregarded by exactly the political actors it was designed to constrain[40]. A regime that regulates the artefact's label rather than the underlying conduct of creating and disseminating a harmful synthetic likeness will always be one offshore tool, or one platform's compression algorithm, away from irrelevance.

6. Comparative Lessons: What Criminal Statutes Elsewhere Add That India Lacks

6.1 Virginia, Texas and the Federal TAKE IT DOWN Act: Criminalising Harm, Not Just Provenance

[39] See Ronin Legal Consulting, India's Draft Deepfake Rules: MeitY's First Regulation Move (2026), https://roninlegalconsulting.com/india-finally-stands-tentatively-up-to-ai-deepfakes/.

[40] "How Europe's Far Right Used Unlabelled AI to Win Votes, and Now Writes the Rules' EUOBSERVER (22 January 2026) <https://euobserver.com/4811/how-europes-far-right-used-unlabelled-ai-to-win-votes-and-now-writes-the-rules/> accessed 13 August 2026.

The comparative contrast this article needs is supplied by jurisdictions that chose to criminalise the harm directly rather than regulate the artefact's label. Virginia amended its existing revenge-pornography statute in 2019 to bring falsely created deepfake images and videos within its reach, making it a Class 1 misdemeanour to disseminate or sell a synthetic nude or sexually explicit depiction of a person without authorisation and with intent to coerce, harass, or intimidate[41]. Texas followed in 2023 with Senate Bill 1361, codified at Section 21.165 of the Texas Penal Code, making it a Class A misdemeanour to knowingly produce or distribute, without consent, a deepfake video depicting a real, identifiable person engaged in sexual conduct or with intimate parts exposed[42], and Texas had separately criminalised deepfake videos of political candidates intended to injure a candidate or influence an election[43]. At the federal level, the TAKE IT DOWN Act, signed into law on 19 May 2025 and in force since that date, makes the intentional online publication of a non-consensual intimate visual depiction of an identifiable person, whether authentic or an AI-generated "digital forgery," a federal crime, and separately criminalises threats to publish such material, while imposing a forty-eight-hour notice-and-removal obligation on covered platforms[44].South Korea supplies a fourth, still more direct comparator. Its 2024 amendment to the legislation governing sexual offences criminalises not only the production and distribution of sexually explicit deepfakes, with a maximum sentence of seven years' imprisonment, but also their mere possession or viewing, and did so by removing the pre-existing

[41]VA. CODE ANN. § 18.2-386.2 (2019), as amended by H.B. 2678, 2019 Va. Acts (adding falsely created videographic or still images to the existing nonconsensual-pornography offence).

[42]TEX. PENAL CODE ANN. § 21.165 (West 2023) (enacted by S.B. 1361, 88th Leg., Reg. Sess. (Tex. 2023), effective Sept. 1, 2023).

[43]See The Current Landscape of Deepfake Legislation in the United States: Analysis of State-Level Responses, 15 J. INFO. POL'Y 97, 100–103 (2025) (Valentine Ugwuoke & Madelyn Rose Sanfilippo).

[44]TAKE IT DOWN Act, Pub. L. No. 119-12, §§ 2–3 (2025); see Is Deepfake Pornography Illegal? Federal and State Laws, CRIMINALDEFENSELAWYER.COM, https://www.criminaldefenselawyer.com/resources/is-deepfake-pornography-illegal.html (last visited Aug. 13, 2026); see also Nonconsensual Deepfake Laws by State: 2026 Tracker, MULTISTATE.AI, https://www.multistate.ai/nonconsensual-deepfake-laws (last visited Aug. 13, 2026).

requirement that prosecutors prove an intent to distribute. That reform goes beyond even the Virginia and Texas statutes discussed below, which remain tied to distribution or dissemination, and illustrates a further point on the spectrum of criminalisation options available to a legislature willing to treat non-consensual synthetic sexual imagery as harmful in itself rather than only upon publication[45]. All three instruments share a common architecture that India's disclosure-focused IT Rules do not attempt: each criminalises the act of creating or distributing specific, harmful synthetic content, rather than requiring that the content, once created, merely be labelled.

6.2 What Transfers to India and What Does Not

The comparative lesson transfers with appropriate caution rather than as a template for direct transplantation. The Virginia and Texas statutes are narrowly targeted, at sexual deepfakes in Virginia's case and at both sexual and electoral deepfakes in Texas's, and both were enacted incrementally, by amending an existing revenge-pornography or election-integrity statute, rather than by creating a single omnibus deepfake offence covering every conceivable synthetic-media harm at once[46]. That incremental, narrowly targeted drafting method is the feature India's proposed BNS and IT Act amendments should follow, for a reason specific to India's constitutional context: narrowly targeted amendments, each keyed to a specific, identifiable harm such as non-consensual sexual imagery or electoral disinformation, are considerably easier to cabin against parody, satire, and political commentary than a single broad standalone deepfake statute would be, precisely because a narrow amendment can be drafted with an explicit carve-out for the specific expressive category it is not meant to reach. A broad, undifferentiated offence covering all 'deceptive synthetic

[45]'South Korea Sexual Deepfake Law' RESEMBLE AI <https://www.resemble.ai/laws-and-regulations/south-korea-sexual-violence-crimes-punishment-act-amendments> (last visited 13 August 2026); see also 'Cheap AI Tools Fuel Teen-Driven Rise in Deepfake Sex Crimes in South Korea' KOREA HERALD (17 November 2025) <https://www.koreaherald.com/article/10616925> accessed 13 August 2026.

[46]See VA. CODE ANN. § 18.2-386.2 (2019); TEX. PENAL CODE ANN. § 21.165 (West 2023).

media' risks capturing political satire of the kind that has already drawn constitutional challenge under India's IT Rules more generally, and would invite exactly the vagueness objection that felled Section 66A of the IT Act in Shreya Singhal. The United States' fragmented, amendment-by-amendment approach, whatever its costs in national coherence, offers India a more constitutionally defensible drafting method than a single sweeping statute would.

7. The Legislative Core the Audits Demand

This article concludes that India's existing legal framework is fundamentally insufficient to address deepfake crimes. It relies on a legally exhausted strategy of executive advisories, ill-fitting civil remedies, and an outdated intermediary liability regime that together leave a substantial accountability gap at the centre of the criminal justice system. The article has built this conclusion through four sequential doctrinal audits.

First, Section 2 demonstrates the statutory void. Applying traditional criminal defamation, obscenity, or identity-theft standards under the BNS and the IT Act to generative AI fails because existing law lacks the doctrinal vocabulary to assign mens rea to algorithmic composition, forcing the state to rely on legally weightless MeitY advisories that, on the reasoning of *Indian Express Newspapers* and *Confederation of Ex-Servicemen Associations*, cannot lawfully create penal obligations. Second, Section 3 exposes the civil-law band-aid: the High Courts have used personality rights and privacy doctrine to grant interim injunctions against synthetic endorsement fraud and voice cloning in *Anil Kapoor v. Simply Life India* and *Arijit Singh v. Codible Ventures*, but this privatised remedy serves celebrities and fails the ordinary victim of non-consensual synthetic pornography, who has no commercial persona to invoke and no comparable route to relief.

Third, Section 4 argues that Section 79 safe harbour structurally fails for deepfakes, because a platform offering a generative model is not merely hosting third-party information but composing the illegal media itself, a distinction the government's own function-based test now concedes; this necessitates a shift from a binary strict-liability-or-immunity framework to a reasonable-care standard assessed at the point of model design. Finally, Section 5 addresses the limits of the 2026 synthetic-media amendments to the IT Rules: while disclosure, labelling, and watermarking are constitutionally sound tools for managing misinformation without chilling speech, and the EU AI Act's Article 50 shows that a comparable transparency model can be built without criminalisation, these tools regulate the artefact's provenance rather than its criminal harm, and remain brittle against offshore circumvention, exactly as European experience with unlabelled election-period deepfakes has already shown.

This article's central diagnosis has since received an unusually direct form of validation from India's own judiciary. On 29 July 2026, a Supreme Court bench comprising the Chief Justice of India and Justices Joymalya Bagchi and V. Mohana, hearing a suo motu writ petition on digital-arrest fraud, called on Parliament to enact dedicated criminal legislation covering both digital-arrest fraud and AI-generated deepfakes, with Justice Bagchi observing that the court itself cannot define a new offence and that existing provisions leave an accused free to contest criminal intent. Solicitor General Tushar Mehta confirmed to the bench that a draft bill addressing both digital arrests and deepfakes was under preparation. That intervention, arriving from the country's highest court rather than from academic commentary, corroborates this article's central claim that the

existing statutory and regulatory architecture is not merely imperfect but insufficient, and that the legislative reform proposed below is no longer only a scholarly recommendation[47].

The article's central prescription, drawn from the comparative audit in Section 6, is a targeted legislative amendment to the BNS and the IT Act that establishes a specific criminal offence for the malicious creation and distribution of materially deceptive synthetic media intended to cause dignitary or financial harm, redefines intermediary liability for generative systems along the reasonable-care lines developed in Section 4.2, and codifies robust speech safeguards for parody, satire, and art modelled on Article 50(4) of the EU AI Act. It should be drafted incrementally, amending the BNS's existing cheating, defamation, and obscenity provisions and the IT Act's existing intermediary framework, on the model of the Virginia and Texas statutes, rather than as a single sweeping standalone offence. This would end the era of ad hoc governance by advisory and replace instruments that cannot bind with a criminal and regulatory core that can.

A Note on Sources and Currency

This article reflects the statutory, regulatory, and litigation record as it stood on 13 August 2026. Two points warrant particular note. First, Section 2.2's treatment of the distinction between prompter mens rea and developer design liability was developed from first principles of BNS doctrine cross-checked against product-liability analysis, since no dedicated Indian secondary source disaggregating the two was located during research; readers working in this area are invited to test that synthesis against subsequent scholarship. Second, the safe-harbour dispute discussed

[47] 'Supreme Court Calls For Separate Offence On Digital Arrests, Harsher Punishment Against Accused' BAR & BENCH (29 July 2026) <https://www.barandbench.com/news/supreme-court-calls-for-separate-offence-on-digital-arrests-harsher-punishment-against-accused> accessed 13 August 2026; see also 'India Supreme Court Demands Criminal Law for Digital Arrest, Deepfake Fraud' TECHTIMES (29 July 2026) <https://www.techtimes.com/articles/321946/20260729/india-supreme-court-demands-criminal-law-digital-arrest-deepfake-fraud.htm> accessed 13 August 2026.

in Section 4.1, *Nitin Jairam Gadkari v. Meta Platforms, Inc.*, was at an active and rapidly developing procedural stage at the time of writing, with the Bombay High Court's most recent order dated 5 August 2026; the litigation should be checked for subsequent developments before this article's account of it is relied upon.