

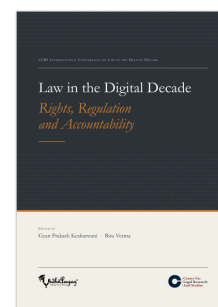
CHAPTER 22

Deepfakes, Defamation, and the Limits of Section 79: A Post-Visakha Analysis in Light of the IT Amendment Rules, 2026

Raiha Owais and Anha Tahir

in *Law in the Digital Decade: Rights, Regulation and Accountability*, edited by Gyan Prakash Kesharwani and Ritu Verma (VidhiAagaz 2026), pp. 267–278

DOI <https://doi.org/10.63108/VAB.LDD.1.22>



CITATIONS

BLUEBOOK (21ST ED.)

Raiha Owais & Anha Tahir, *Deepfakes, Defamation, and the Limits of Section 79: A Post-Visakha Analysis in Light of the IT Amendment Rules, 2026*, in *LAW IN THE DIGITAL DECADE: RIGHTS, REGULATION AND ACCOUNTABILITY* 267 (Gyan Prakash Kesharwani & Ritu Verma eds., VidhiAagaz 2026).

OSCOLA (4TH ED.)

Raiha Owais and Anha Tahir, ‘Deepfakes, Defamation, and the Limits of Section 79: A Post-Visakha Analysis in Light of the IT Amendment Rules, 2026’ in Gyan Prakash Kesharwani and Ritu Verma (eds), *Law in the Digital Decade: Rights, Regulation and Accountability* (VidhiAagaz 2026) 267.

CHICAGO (17TH ED.)

Owais, Raiha, and Anha Tahir. “Deepfakes, Defamation, and the Limits of Section 79: A Post-Visakha Analysis in Light of the IT Amendment Rules, 2026.” In *Law in the Digital Decade: Rights, Regulation and Accountability*, edited by Gyan Prakash Kesharwani and Ritu Verma, 267–278. Maihar: VidhiAagaz, 2026. <https://doi.org/10.63108/VAB.LDD.1.22>.

APA (7TH ED.)

Owais, R., & Tahir, A. (2026). Deepfakes, defamation, and the limits of Section 79: A post-visakha analysis in light of the IT amendment rules, 2026. In G. P. Kesharwani & R. Verma (Eds.), *Law in the digital decade: Rights, regulation and accountability* (pp. 267–278). VidhiAagaz. <https://doi.org/10.63108/VAB.LDD.1.22>

SOURCE

<https://doi.org/10.63108/VAB.LDD.1.22>

LICENCE

Open access under the Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0): it may be copied, shared and adapted for non-commercial purposes, with credit to the authors and the source.

Citations are given as a guide; check them against the style manual you follow. This page was added by VidhiAagaz to the downloaded file. The chapter follows, with the page numbers of the printed book.

CHAPTER 22

Deepfakes, Defamation, and the Limits of Section 79: A Post-Visakha Analysis in Light of the IT Amendment Rules, 2026

RAIHA OWAIS* AND ANHA TAHIR**

The introduction of deepfake technology has long exposed gaps in India's intermediary liability system under Section 79 of the Information Technology Act, 2000. This paper revisits Google India Pvt. Ltd. v. Visakha Industries Ltd., which clarified intermediary liability for defamatory content through the "publication by inaction" doctrine, holding that an intermediary's failure to act on actual knowledge of unlawful content may defeat the statutory safe harbour. While Visakha addressed static, user-generated defamation, this paper argues that its reasoning does not sit well with synthetically generated deepfake content, in which the intermediary's own algorithmic infrastructure is implicated in fabrication rather than mere hosting.

The recent Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, notified 10 February 2026 and enforced from 20 February, represents the legislature's first direct response to this gap, introducing a formal definition of "Synthetically Generated Information" (SGI), mandatory labelling and metadata-tracing obligations, and a 3-hour takedown window for deepfake content, with loss of safe harbour for non-compliant intermediaries. This paper examines whether the 2026 Amendment adequately closes the doctrinal fissure left open by Visakha, or whether its "actual knowledge" and verification standards remain too narrow to address content that is synthetically fabricated rather than merely spread.

Moreover, this paper considers the relationship between the 2026 Amendment, the pending Draft Second Amendment Rules of March 2026, and emerging criminal liability under the Bharatiya Nyaya Sanhita, 2023, for defamation and identity-based offences. It concludes by proposing a graduated due diligence standard tailored to SGI, arguing that labelling and takedown timelines alone cannot substitute for a coherent liability framework that distinguishes conduit from co-creator.

Keywords: Intermediary Liability, Deepfakes, Section 79 IT Act, Synthetically Generated Information (SGI), Safe Harbour

1 INTRODUCTION

The technology behind deepfakes is the output of two competing neural networks, a generator that creates fake content and a discriminator that tries to detect it, trained against each other until the generator's output becomes indistinguishable from an authentic recording.¹ This technique, first demonstrated in 2014 and popularised under the "deepfake" label from 2017 onward, has since migrated from research laboratories into consumer-facing applications capable of producing a photorealistic video or a cloned voice from a handful of source images or seconds of audio.²

* Faculty of Law, Aligarh Muslim University, Aligarh, Uttar Pradesh, India.

** Faculty of Law, Aligarh Muslim University, Aligarh, Uttar Pradesh, India.

¹ Kashish Gupta, "THE FUTURE OF DEEPFAKES: NEED FOR REGULATION" 5 National Law University Delhi (2023)

² John Thayer, 'Defamation or Impersonation? Working Towards a Legislative Remedy for Deepfake Election Misinformation', 66 *William & Mary Law Review* 255 (2024).

The damage caused by it is immense. In 2018, the investigative journalist Rana Ayyub discovered that a pornographic video bearing her face, generated from publicly available photographs, was circulating on WhatsApp and Twitter, reportedly viewed and shared multiple times within days.³ Unlike a forged photograph pasted onto a bulletin board, in Ayyub's case this was defamatory and sexually humiliating, and an algorithm manufactured the statement made about her.

India's principal statutory answer to intermediary liability for unlawful third-party content, Section 79 of the Information Technology Act, 2000 ('the IT Act'), was not written with this problem in mind. It was drafted for a much older internet world of static user content: something a real, identifiable person actually wrote, with the platform just hosting or passing it along. The most authoritative judicial gloss on that architecture, the Supreme Court's 2019 decision in *Google India Private Ltd. v. M/S Visakha Industries*⁴, addressed such a scenario. An allegedly defamatory article was posted by one identified individual on a Google Group, which the intermediary declined to remove despite notice. The Court's resulting "publication by inaction" doctrine, borrowed from English authorities on club noticeboards and internet service providers, held that an intermediary's continued, knowing inaction in the face of a removable defamatory posting could itself constitute "publication" for criminal defamation under the Indian Penal Code.

This paper asks whether that doctrine, forged for passive hosting of human-authored text, can meaningfully govern a technology in which the intermediary's own generative infrastructure participates in the making of the defamatory statement. It argues that it cannot do so without strain, because *Visakha* assumes three things that deepfakes break: a fixed, identifiable piece of content; one clear moment when the platform "knew"; and a simple choice between taking something down or leaving it up. But when the platform itself provides the tool (a face-swap filter, a voice cloner, an image generator), the real question isn't whether it failed to remove something after being notified. It's whether it helped make the thing in the first place.

2 SECTION 79 AND THE ARCHITECTURE OF INTERMEDIARY IMMUNITY

Section 79, before its 2009 substitution, provided the narrower shield as it exempted what the statute then called a network service provider from liability under the IT Act and its subordinate rules or regulations for third-party information it made available, but only if the provider could show both that it had no knowledge of the offending act and that it had exercised due diligence to prevent it.⁵ As the Supreme Court later clarified in *Visakha*, this pre-2009 exemption was narrow in a specific sense, because it sat in Chapter XII of the IT Act, it shielded intermediaries only from liability arising under the Act itself, and did not touch offences under general criminal law such as defamation under Section 356 of the Bharatiya Nyaya Sanhita (previously Sections 499–501 of the Indian Penal Code).⁶

The 2009 amendment changed this significantly. The revised provision opens with a non-obstante clause overriding conflicting law, exempting intermediaries from liability for third-party content they host, language broad enough to cover general criminal law, unlike the earlier version. This immunity is conditional: under Section 79(3)(b), it lapses if the intermediary fails to act expeditiously after gaining actual

³ Abigail George, 'Defamation in the Time of Deepfakes', 45(1) Columbia Journal of Gender and Law 45 (2024) (recounting Rana Ayyub, 'I Was the Victim of a Deepfake Porn Plot Intended to Silence Me', HuffPost, 21 November 2018).

⁴ AIR 2020 SC 350.

⁵ *Supra* note 4.

⁶ *Ibid*

knowledge or receiving government notification of unlawful content. Under the 2011 Intermediary Guidelines Rules, Rule 3(4) required takedown once the intermediary knew, or was told, that hosted material was defamatory (among other things). Both these provisions were challenged before the Supreme Court in *Shreya Singhal v. Union of India*, and the Court, concerned about turning intermediaries into unilateral judges over the legality of the vast volume of content that passes through them, narrowed the meaning of actual knowledge in Section 79(3)(b) and Rule 3(4) so that it arises only once a court order or an appropriate-government notification communicates it.⁷ *Visakha* later summed up the reasoning as essentially practical: without that kind of external, authoritative trigger, large intermediaries would have no real way to sort the genuinely legitimate takedown demands from the flood of requests they receive every day.⁸ Constitutionally, then, there is safe harbour immunity from general law, and not merely from liability under the IT Act, at the cost of a narrow, externally-gated knowledge standard.

The 2021 Intermediary Guidelines and Digital Media Ethics Code Rules layered further due-diligence obligations onto this structure, particularly for larger platforms designated as significant social media intermediaries: a grievance redressal mechanism, expeditious removal of flagged sexually explicit or impersonating content, and encouragement, though not an unconditional mandate, to deploy automated tools against certain categories of unlawful material, subject to periodic human review.⁹ Even here, the underlying conception of the intermediary's role did not change: it remained that of a facilitator with no editorial control over content authored by others, a characterisation the Government of India itself pressed in *Visakha* when distinguishing intermediaries from newspaper editors under the Press and Registration of Books Act, 1867.¹⁰

This entire architecture rests on the statutory definition of an intermediary under Section 2(1)(w) of the IT Act as an entity that, on behalf of another person, receives, stores or transmits an electronic record, or otherwise provides a service in respect of it. As the Government of India's own AI Governance Guidelines Committee has since observed, a definition built around telecom carriers, search engines and cyber cafes sits uneasily with AI systems that produce output from user prompts, or even act autonomously, refining their results through continuous learning.¹¹ The Committee went further, noting that Section 79 immunity depends on the intermediary not initiating transmission, choosing the recipient, or altering the data in question, conditions that many content-generating AI systems may simply fail to satisfy.¹²

3 THE PUBLICATION-BY-INACTION DOCTRINE

The complainant company initiated criminal defamation proceedings against accused 1, the coordinator of “Ban Asbestos Group”, for authoring and publishing allegedly defamatory articles titled “Visakha Asbestos Industries making gains” (on 31-07-2008) and “Poisoning the system” (on 20-11-2008), and accused 2 (Appellant), Google India Pvt. Ltd., for hosting the publications. On 09-12-2008, the complainant issued a notice for takedown; the appellant failed to do so. Google India resisted on several

⁷ *Shreya Singhal* (2015) 5 SCC 1.

⁸ *Supra* note 4.

⁹ *Supra* note 1.

¹⁰ *Gambhirsinh R. Dekare v. Falgunbhai Chimabhai Patel*.

¹¹ Government of India, India AI Governance Guidelines (Ministry of Electronics and Information Technology, 5 November 2025).

¹² *Supra* note 11.

grounds: that it was not itself the intermediary (Google LLC, its US parent, owned and operated Google Groups), that the notice did not specify the objectionable URL and that, as a passive host, it could not be attributed actual knowledge of content posted by millions of users without editorial oversight.¹³

Drawing from *Byrne v. Deane* (1937) and *Godfrey v. Demon Internet Ltd.* (2001), the Court held that the test for publication in an internet context is that a party which, having the power, the right and the ability to remove the defamatory matter, and being called upon to do so, defies and rejects the request may itself be treated as a publisher. Passive hosting alone does not attract liability; it is the deliberate inaction after notice that does. However, for the offence under Section 500 IPC (Section 356(2) BNS), publication alone is insufficient. The accused must also have acted with the requisite intention or knowledge that the content would harm the reputation of the person concerned. This meant that an intermediary who is alerted to defamatory material and possesses the technical means to take it down, but refuses to do so, may thereby become a publisher of it, provided the separate mental element required by Section 356, intention or knowledge of likely harm to reputation, is also independently made out.¹⁴

Two qualifications the Court itself attached to this holding deserve emphasis, because they mark the doctrine's outer edge. First, the Court expressly noted that the noticeboard analogy from *Byrne* is not straightforwardly transferable to the internet, observing that the scale and structure of online publication are materially different from a wall in a golf club.¹⁵ Second, and more importantly for present purposes, the Court left open whether, absent a court order, an intermediary could lawfully insist on judicial adjudication before removing content even where it was, on the facts, the true intermediary. The eventual disposition was correspondingly narrow: the Court set aside the High Court's findings on the merits of Google India's refusal to act on notice, and remanded the matter to the trial Magistrate to determine, on evidence, whether the appellant was in fact the intermediary and whether publication had occurred.¹⁶

What *Visakha* decided, in other words, was a question about the legal consequence of inaction toward a fixed, human-authored object of publication whose only relevant lifecycle event before the court was a single act of hosting and a single, later act of refusal to de-list. The appellant's conduct at issue was entirely custodial: it did not write, edit, or algorithmically transform the Ban Asbestos India articles; its only alleged wrongdoing was declining to take them down. That is precisely the fact pattern for which the notice-and-knowledge architecture of Section 79, as read down in *Shreya Singhal*¹⁷, was designed.

4 FROM CONDUIT TO CO-CREATOR: WHY VISAKHA DOES NOT TRAVEL WELL TO SYNTHETIC MEDIA

Deepfakes disturb the *Visakha* framework at each of the three junctures identified above: the determinacy of the object of publication, the meaningfulness of a single point of "actual knowledge," and the binary remove-or-retain choice on which the publication-by-inaction test depends.

On the first point, a defamatory article is, once posted, a fixed textual artefact: the same words persist until edited or removed, and "the defamatory matter" referred to throughout the *Visakha* judgment is a stable referent. A deepfake generated through a generative adversarial network, by contrast, is the terminal

¹³ *Supra* note 4.

¹⁴ *Supra* note 4.

¹⁵ *Supra* note 4.

¹⁶ *Ibid*

¹⁷ (2015) 5 SCC 1.

output of an iterative training process in which a generator network continuously refines its forgeries against a discriminator network until the two become computationally indistinguishable.¹⁸ Because the underlying model can regenerate near-identical but technically distinct outputs on demand, and because photorealistic synthetic media collapses, in a way ordinary text does not, the reader's ability to distinguish signifier from signified, a deepfake's capacity to cause reputational harm survives even the viewer's knowledge that it is fake.¹⁹ The Byrne noticeboard test asks, in essence, whether the defendant's own conduct effectively kept defamatory matter in place; but where the matter in question can be trivially regenerated, subtly altered to defeat hash-based detection, and redistributed through parallel channels, the very idea of one fixed item whose ongoing presence a defendant can be said to have caused becomes strained.

On the second point, *Visakha's* entire notice-based architecture presumes an identifiable moment at which knowledge attaches, the 9 December 2008 letter in that case, after which the intermediary's continued inaction becomes legally significant. Rana Ayyub's experience illustrates that the video was not merely disseminated widely but reproduced and re-shared across WhatsApp and Twitter simultaneously and virally, with no single upload event or single platform from which a notice-and-takedown obligation could meaningfully flow.²⁰ The Supreme Court itself, in *Shreya Singhal*, narrowed the actual-knowledge standard, reasoning that requiring intermediaries to adjudicate legitimacy across the volume of the internet, with countless posts every minute even back in 2015, risked either overwhelming platforms or effectively handing them a private censorship role that the Court was unwilling to sanction. Generative tools multiply this volume problem by an additional order of magnitude; content moderation research cited in the comparative deepfake-regulation literature confirms that automated detection tools deployed against synthetic media are themselves imperfect and even good-faith compliance with a notice may not reliably identify or suppress all derivative copies.²¹

The third and most consequential disturbance concerns the remove-or-retain binary itself. In *Visakha*, the intermediary's only relevant choice, when notified, was whether to take down content that a third party had already, independently, authored. Where the offending content is instead produced using a generative feature that the intermediary itself built into its platform, a generator function, the intermediary's causally significant act occurs before any question of notice arises at all: it lies in having supplied the generative capability that made the fabrication possible. This is not a hosting decision at all; it more closely resembles the kind of conduct the AI Governance Guidelines Committee flagged as falling outside the immunity altogether, initiating a transmission, choosing a recipient, or altering the underlying data, conduct that Section 79 was never designed to cover, and which the Committee concluded may put many generative AI systems beyond the safe harbour independent of any subsequent takedown failure.²² Put differently, *Visakha's* publication-by-inaction doctrine is built to answer whether an intermediary wrongly kept hosting something it should have taken down. Deepfakes produced through platform-native tools instead raise the logically prior question of whether the intermediary wrongly helped make the thing in

¹⁸ *Supra* note 2 at 256.

¹⁹ Mohit Kar and Shreya Sahoo, "DEEPFAKES AND ITS INIQUITIES: REGULATING THE DARK SIDE OF AI" 5.1 *National Law University of Odisha Student Law Journal* 58 (2020)

²⁰ *Supra* note 3

²¹ Andrew Ray, 'Disinformation, Deepfakes and Democracies: The Need for Legislative Reform', 44(3) *UNSW Law Journal* 983 (2021).

²² *Supra* note 11

the first place, a question closer to authorship or origination under the IT Act's own definitions than to custody, and one the publication-by-inaction test was never built to ask.

This is not merely a theoretical distinction. Comparative commentary on deepfake-enabled election misinformation in the United States has identified an analogous structural gap in Section 230 of the Communications Decency Act, which likewise assumes a sharp separation between the platform and the third-party source of the content it hosts, a separation that erodes once platforms themselves supply generative tools that materially contribute to the content's creation.²³ Free-speech scholarship on synthetic media has similarly observed that existing doctrinal exceptions built for authored falsehoods (defamation, intentional infliction of emotional distress, and the right of publicity) were not designed with an algorithmic co-author in mind, and each must be expanded to reach a wrong whose proximate cause is a model rather than a person.²⁴ The right-of-publicity literature makes a parallel point from a different angle: that doctrine emerged historically to address the portable camera and mass media together, as a *combination* of a new way of capturing someone's likeness and a new way of disseminating it, and deepfakes present exactly that same two-part structure, except that the device doing the capturing is now owned and operated, in many cases, by the very platform accused of merely hosting the result.²⁵

None of this suggests that *Visakha* was wrongly decided on its own facts, nor that the publication-by-inaction doctrine is without continuing value. For content that remains genuinely third-party-authored and merely hosted, such as a deepfake created with an independent, off-platform tool and later uploaded to a hosting service, *Visakha*'s notice-based inquiry remains apt and indeed necessary, since it is the mechanism by which victims can compel removal without waiting for a court order in every case where the intermediary is truly custodial. The argument, rather, is that *Visakha* cannot be the *sole* doctrinal resource for deepfake-enabled defamation, because a meaningful share of real-world deepfake harm is produced using generative infrastructure that the hosting platform itself supplies, and for that category the publication-by-inaction inquiry is simply the wrong question.

5 THE IT AMENDMENT RULES, 2026: DOES THE LEGISLATURE CLOSE THE GAP?

MeitY notified the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 by Gazette Notification G.S.R. 120(E) on 10 February 2026. These amendments came into force on 20 February 2026. The Amendment is India's first substantive statutory framework directed at AI-generated content as such, following an October 2025 draft that had circulated for stakeholder consultation, as several contemporaneous commentators have noted.²⁶

The centrepiece of the 2026 Amendment is a new, technology-neutral definition of "Synthetically Generated Information" (SGI) which is audio, visual, or audio-visual content that has been created or algorithmically altered using computer resources in a manner that makes it appear authentic, such that a person would be misled into believing it to be genuine.²⁷ Two features should be flagged. First, it is not confined to AI-generated content in the narrow, model-based sense: it is deliberately broad enough to

²³ *Supra* note 2 at 277

²⁴ Jacob Bourgault, 'Free Speech and Synthetic Lies: Deepfakes, Synthetic Media, and the First Amendment', 3(1) *Journal of Intellectual Property Law*.

²⁵ Michael P Goodyear, "Dignity and Deepfakes" 57 *Arizona State Law Journal* 931 (2025).

²⁶ AZB & Partners, 'India's New Law Governing Synthetic Media' available at, <https://www.azbpartners.com/bank/88502/> (last visited on August 28, 2026).

²⁷ The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, r 2(wa)

capture other forms of algorithmic alteration as well, a choice several commentators regard as inviting both over- and under-inclusion.²⁸ Further, the Rules impose their obligations on the *intermediary that hosts* SGI, largely without differentiating between an intermediary that merely stores synthetic content uploaded from elsewhere and one whose own in-platform tools generated it.

Substantively, the 2026 Amendment layers three obligations onto covered intermediaries. First, mandatory labelling: SGI must carry a label that is prominent and noticeable to an ordinary viewer or listener, a qualitative standard that replaced an earlier, more mechanical draft proposal requiring a watermark covering at least ten per cent of a visual work's surface area, or running through the first ten per cent of an audio clip's duration.²⁹ Second, permanent, non-strippable provenance metadata, broadly consonant with international content-authentication standards such as the Coalition for Content Provenance and Authenticity (C2PA), which the Government's own AI Governance Guidelines Committee had recommended examining as a mechanism for tracing whether content was AI-generated or modified.³⁰ Third, and most sharply, a compressed takedown window reported as three hours for flagged unlawful synthetic content generally, a substantial reduction from the up-to-thirty-six-hour windows the 2021 Rules allowed for government or court-ordered takedowns, with an even shorter window reported for non-consensual sexual imagery specifically.³¹ Non-compliance carries a severe consequence: forfeiture of Section 79 safe harbour, with the non-compliant intermediary potentially treated, for liability purposes, as if it had itself created the unlawful content.³²

The doctrinal issue this paper identifies survives the 2026 Amendment largely intact, for four related reasons. First, the safe-harbour forfeiture mechanism is still triggered by a failure to act within the takedown window following a complaint or governmental notification. In other words, it retains precisely the “*actual knowledge*” gate that Shreya Singhal read into Section 79(3)(b) and that *Visakha* applied to a purely custodial host.³³ The 2026 Amendment compresses the clock; it does not relocate the trigger. It therefore leaves undisturbed the prior question this paper has pressed, whether the intermediary's own generative infrastructure contributed to the fabrication, which remains analytically anterior to, and un-addressed by, any timeline for post-notice removal.

Second, the Internet Freedom Foundation's detailed submissions on the preceding draft, several of which persisted into critique of the final Rules, warned that the SGI definition's reliance on whether content would appear authentic to a reasonable viewer is broad and subjective, sweeping in low-risk domestic uses of AI, filters, translation tools, alongside high-risk political or sexual deepfakes, without any risk-tiering to differentiate intermediary obligations accordingly.³⁴ A regime that regulates by appearance rather than by the intermediary's own causal proximity to the generative act cannot, by construction, distinguish a platform that merely hosts an externally-produced deepfake from one whose own tools produced it, the very distinction this paper argues is doctrinally essential.

²⁸ ANM Global, ‘Regulating Synthetic Content: Key Implications of the IT Rules Amendment, 2026’ *available at*, <https://anmgloba.net/articles/regulating-synthetic-content-key-implications-of-the-it-rules-amendment-2026> (last visited on August 29, 2026).

²⁹ *Supra* note 28.

³⁰ *Supra* note 11.

³¹ *Supra* note 26.

³² The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, r 7; *supra* note 26.

³³ *Supra* note 4.

³⁴ Internet Freedom Foundation, ‘IT Intermediary Amendment Rules, 2026 Contradict Their Purpose’ *available at*, <https://internet-freedom.in/> (last visited on August 27, 2026).

Third, constitutional commentary on the notified Rules has raised concerns continuous with those that led the Supreme Court to read down Section 79(3)(b) in *Shreya Singhal* in the first place. One detailed critique argues that mandatory content-based labelling risks operating as compelled speech, and that provisions conditioning safe harbour on compliance with MeitY's clarifications, advisories and directions, carried forward and expanded in the subsequent Draft Second Amendment, may reproduce the vagueness and overbreadth for which Section 66A of the IT Act was struck down in *Shreya Singhal*.³⁵ To the extent these objections are well-founded, they counsel against simply widening the existing notice-and-takedown gate further (for instance, by shortening timelines even more or broadening the SGI definition still further), since doing so intensifies the very features of the pre-existing architecture that this paper argues are ill-suited to the generation-stage problem, rather than curing them.

Fourth, and practically, the compressed three-hour window is reported to strain even well-resourced intermediaries' capacity to verify both the *legality* of a takedown request and the authenticity of the underlying content (that is, whether it is synthetic at all, and whether provenance metadata has been stripped or spoofed) within the time allowed, a strain analysts have suggested will incentivise precautionary, over-inclusive removals with attendant costs to legitimate speech.³⁶ This is, in effect, the volume-and-verification problem the Supreme Court worried about in *Shreya Singhal* and *Visakha*, replicated in a harder register: intermediaries are now expected not merely to judge legality quickly, but to judge legality *and* technical authenticity quickly, using a legal test (actual knowledge on notice) that was calibrated for neither task.

The net assessment, then, is that the 2026 Amendment answers the question of how to make synthetic content identifiable and remove it quickly once flagged, but leaves substantially open the question this paper considers doctrinally prior: who is responsible, independent of any notice, when the platform's own infrastructure was used to fabricate the content in the first place. Because *Visakha*'s publication-by-inaction test cannot answer that question, and because the 2026 Amendment's safe-harbour forfeiture mechanism operates through the same notice-triggered gate *Visakha* applied, the doctrinal fissure between conduit and co-creator persists after the 2026 Amendment, notwithstanding its genuine advances on labelling and provenance.

6 THE DRAFT SECOND AMENDMENT RULES, MARCH 2026, AND CRIMINAL LIABILITY UNDER THE BHARATIYA NYAYA SANHITA

On 30 March 2026, MeitY published the Draft Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Second Amendment Rules, 2026 for public consultation, issued under Sections 69A, 79 and 87 of the IT Act. Whereas the February 2026 Amendment focused on SGI as a content category, the Draft Second Amendment shifts the register toward institutional and user-level obligations. It proposes to extend publisher-like duties (grievance-officer contact details, code-of-ethics adherence, and fixed takedown responsiveness) to individual users who post "news and current affairs" content beyond prescribed thresholds, treating them, in effect, as publishers; it proposes a continuous, duration-long display requirement for SGI labels rather than a momentary disclosure; and it proposes an inter-

³⁵ Rudraksh Lakra, 'Guest Post: A Constitutional Critique of the Synthetically Generated Information (IT Rules Amendment), 2026', Indian Constitutional Law and Philosophy (blog, 11 February 2026).

³⁶ *Supra* note 28

ministerial “Digital Content Review Panel” combining MeitY and Ministry of Information and Broadcasting officials to adjudicate user–publisher disputes.³⁷

For the purposes of this paper’s argument, the Draft Second Amendment is best read as complementary rather than corrective. Its innovations address a different structural gap (the blurring line between individual social media users and institutional publishers) rather than the conduit/co-creator distinction pressed in Parts 4 and 5. If anything, by directing regulatory and public attention toward user-level publisher obligations and toward compliance with MeitY’s advisory directions, the Draft Second Amendment risks further deferring scrutiny of the intermediary’s own generative-AI features, which remain regulated, if at all, through the same notice-triggered SGI takedown mechanism examined in Part 5.

Running beneath both sets of Rules is the substantive criminal law under the Bharatiya Nyaya Sanhita, 2023, which replaced the Indian Penal Code with effect from 1 July 2024. Section 356 BNS³⁸ replaces IPC Section 500 as the offence of criminal defamation, retaining the core “making or publishing” structure that the *Visakha* Court applied to Google India. Section 336 BNS addresses forgery for harming reputation, what commentators describe as enhanced clarity for digital forgery and impersonation, a provision of obvious relevance to a deepfake that fabricates a statement or image attributed to the victim. Sections 351 and 353 BNS³⁹ address criminal intimidation and statements intended to cause public mischief through digital platforms, Section 77⁴⁰ extends voyeurism-adjacent liability to non-consensual capture and circulation of intimate images, and Section 319⁴¹ addresses cheating by personation, potentially relevant to voice-cloning fraud. The Digital Personal Data Protection Act, 2023 supplies a further, non-criminal layer, since the unauthorised processing of biometric data (a face or a voice) to construct a deepfake independently attracts penalties under that statute, reported as high as ₹250 crore for serious violations.⁴²

Because these BNS offences, like their IPC predecessors in *Visakha*, sit outside Chapter XII of the IT Act, the *Visakha* Court’s foundational holding that Section 79 does not immunise an intermediary from criminal liability under general law once “publication” is independently established continues to apply in principle after the 2026 Amendment. But this is precisely where the conduit/co-creator distinction re-emerges with sharpened stakes: attributing the “making” of a synthetic statement of fact under BNS Section 356 to a platform’s generative model, rather than to an individual human actor, poses an actus reus and mens rea problem that neither the 2026 Amendment nor the Draft Second Amendment purports to resolve, because both instruments operate in the regulatory register of safe-harbour conditions rather than the criminal-attribution register of who “made” the statement. A coherent answer to that criminal-law question, this paper suggests in Part 7, requires the same conduit/co-creator sorting this paper has pressed throughout: an evidentiary sorting mechanism analogous to the rebuttable presumption

³⁷ Government of India Ministry of Electronics and Information Technology, ‘Draft amendments to the IT Rules, 2021 in relation to strengthening intermediary compliance with clarifications, advisories and directions issued by the Ministry and digital media oversight’ available at <https://www.meity.gov.in/documents/act-and-policies/draft-amendments-to-the-it-rules-2021-in-relation-to-strengthening-intermediary-compliance-with-clarifications-advisories-and-directions-issued-by-the-ministry-and-digital-media-oversight-AjM2YjMtQWw?pg=1&title=Draft-amendments-to-the-IT-Rules%2C-2021-in-relation-to-strengthening-intermediary-compliance-with-clarifications%2C-advisories-and-directions-issued-by-the-Ministry-and-digital-media-oversight> (last visited on Sep 21, 2026)

³⁸ Bharatiya Nyaya Sanhita, 45 of 2023, s 356.

³⁹ Bharatiya Nyaya Sanhita, 45 of 2023, ss 351 and 353.

⁴⁰ Bharatiya Nyaya Sanhita, 45 of 2023, s 77.

⁴¹ Bharatiya Nyaya Sanhita, 45 of 2023, s 319.

⁴² The Digital Personal Data Protection Act, 2023 (Act 22 of 2023), The Schedule, item 1.

of editorial responsibility the *Visakha* Court itself discussed in the context of newspaper editors under the Press and Registration of Books Act.⁴³

7 TOWARDS A GRADUATED DUE DILIGENCE STANDARD FOR SYNTHETICALLY GENERATED INFORMATION

The preceding Parts converge on a single proposition: labelling obligations and compressed takedown timelines regulate the *visibility* of synthetic content and the *speed* of the response to a complaint, but neither allocates responsibility according to an intermediary's *causal proximity* to the act of synthesis. That question is precisely the one the Government's own AI Governance Guidelines Committee flagged as unresolved when it called for clarity on how liability should be apportioned "across the AI value chain", developer, deployer, and user, under an intermediary-liability statute never drafted with that chain in mind.⁴⁴ This paper proposes a three-tier due diligence standard responsive to that call, built around the intermediary's relationship to the generative act rather than solely around its response to notice.

Pure Conduit. Where an intermediary merely hosts SGI that was authored and generated entirely by a third-party using tools external to the platform (the closest analogue to the facts in *Visakha* itself), the existing publication-by-inaction, notice-based framework remains appropriate and should continue to govern. Safe harbour should remain available, subject to the labelling, provenance-preservation, and compressed takedown obligations the 2026 Amendment already imposes.⁴⁵ This is the category for which *Visakha*'s doctrine was designed, and it continues to do useful work here.

Enabling Infrastructure. Where an intermediary provides in-platform generative tools that a user then operates to produce the SGI, the intermediary should bear a heightened *proactive* due diligence duty that attaches independently of any subsequent notice. Concretely, this would include mandatory embedding of provenance metadata at the point of generation (operationalising the C2PA-style standards the AI Governance Guidelines Committee has already endorsed for study),⁴⁶ technical safeguards against generating outputs that mimic identifiable real persons without their consent (for instance, matching against opt-out registries for public figures and complainants), and liability for failure to implement regardless of whether a takedown notice was ever served.

Co-creation and Active Curation. Where an intermediary's own algorithmic systems (recommendation engines that substantially amplify synthetic content, or those that synthesise third-party inputs into new outputs) themselves modify or synthesise content rather than merely transmitting it unaltered, the intermediary should be treated, for that specific output, as akin to a publisher or originator, and should not be entitled to Section 79 immunity for it at all. This tier operationalises the AI Governance Guidelines Committee's own observation that Section 79 immunity was conditioned on the intermediary not modifying transmitted data, and that many AI systems fall outside that condition as a matter of ordinary statutory construction, independent of any amendment.⁴⁷ It also extends, into the algorithmic context, the conduit/editor distinction the *Visakha* Court itself drew from the newspaper context, where an editor

⁴³ *Supra* note 10

⁴⁴ *Supra* note 11

⁴⁵ *Supra* note 26.

⁴⁶ *Supra* note 11

⁴⁷ *Ibid*

exercising selection and control over published matter attracts a presumption of responsibility that a mere distributor does not.⁴⁸

Two safeguards would keep this graduated standard within the constitutional boundaries. First, judicial or designated-officer certification should remain the trigger for any *content-based* unlawfulness determination: whether a given piece of synthetic content is in fact defamatory, or otherwise unlawful, should continue to require an external adjudicative gate, consistent with *Shreya Singhal's* read-down of “actual knowledge”.⁴⁹ Second, and critically, the Tier Two obligations proposed above (provenance embedding and consent-matching safeguards) should be delinked entirely from that gate, because they are *ex ante design obligations* rather than *ex post content-adjudication obligations*. A duty to embed a watermark at the point of generation does not require anyone to judge whether the resulting content is defamatory; it only requires the platform to build its tools responsibly. Decoupling these two obligation-types in this manner should substantially answer the overbreadth and vagueness objections the Internet Freedom Foundation and constitutional commentators have raised against provisions that condition safe harbour on broad, content-based labelling duties or on compliance with open-ended ministerial advisories.⁵⁰

Finally, this graduated standard would also assist the criminal-law attribution problem identified in Part 6. By these clear tiers of intermediary proximity to the generative act, it would supply courts adjudicating BNS Section 356 prosecutions with an evidentiary framework analogous to the rebuttable presumption of editorial responsibility already recognised for print editors, allowing a Tier Three intermediary’s algorithmic co-creation to ground an inference of “making” under the BNS, while preserving Tier One intermediaries’ position as mere custodians whose liability continues to depend on proof of knowing inaction, exactly as *Visakha* contemplated.

8 CONCLUSION

Google India Pvt. Ltd. v. Visakha Industries Ltd. filled a genuine gap in Indian intermediary-liability jurisprudence, establishing that an intermediary with both the power and the practical ability to remove defamatory material, which knowingly declines to do so, may be treated as a publisher through its inaction, whatever its ostensibly custodial role.⁵¹ That doctrine remains sound for the case it was built for: a static, human-authored posting merely hosted by a passive conduit.

It does not, however, describe deepfakes produced using an intermediary’s own generative infrastructure. There, the prior question is not whether the intermediary failed to remove something after notice, but whether its tools helped fabricate that thing in the first place, a question of origination that publication-by-inaction was never designed to ask.

The IT Amendment Rules, 2026 mark real progress: India’s first statutory definition of Synthetically Generated Information, mandatory labelling and provenance-tracing, and a compressed takedown window.⁵² Yet because safe-harbour forfeiture still hinges on the same actual-knowledge, notice-triggered gate *Shreya Singhal* read into Section 79(3)(b) and *Visakha* applied to a custodial host, the conduit/co-creator distinction remains largely unresolved, a gap independent critiques of the Rules bear out. The

⁴⁸ *Supra* note 10

⁴⁹ (2015) 5 SCC 1.

⁵⁰ *Supra* note 35.

⁵¹ *Supra* note 4.

⁵² *Supra* note 26.

Draft Second Amendment Rules and the Bharatiya Nyaya Sanhita, 2023 add complementary but non-coextensive layers of publisher and criminal liability.⁵³

A graduated due diligence standard, tying obligations to an intermediary's causal proximity to the act of synthesis rather than only to its response to notice, offers a way forward, one faithful to *Visakha's* own insight that control over content, not merely custody of it, should determine liability.

* * *

⁵³ *Supra* note 39.