VEDANT CHOUDHARY
Associate

ANKIT CHUGH
Intern

India's 2026 Amendment to IT Rules: Regulation of Deepfakes, AI Content and the Three-Hour Takedown Regime

From Regulatory Gap to Regulatory Framework

On 10 February 2026, the Government of India notified the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 (*vide* Gazette notification G.S.R. 120(E)),[1] amending the framework introduced in 2021 under the Information Technology Act, 2000. The amendment, effective from 20 February 2026, represents a significant development in India's digital regulatory sphere by expressly addressing deepfakes and AI-generated content.[2]

The parent framework, the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, was notified on 25 February 2021 (vide G.S.R. 139(E)) under Section 87(2) of the IT Act, superseding the Information Technology (Intermediary Guidelines) Rules, 2011. Administered jointly by the Ministry of Electronics and Information Technology and the Ministry of Information and Broadcasting, the 2021 Rules established a comprehensive due diligence framework for intermediaries, including notice-and-takedown requirements, grievance redressal mechanisms, and a distinction between social media intermediaries and Significant Social Media Intermediaries (**SSMIs**) based on user thresholds. The Rules were subsequently amended on 28 October 2022, 6 April 2023, and 15 November 2025, progressively expanding intermediary accountability. However, none of these iterations explicitly defined or regulated synthetic or AI-created media.

When the 2021 Rules were notified, generative AI tools capable of producing hyper-realistic video, voice cloning, and large-scale synthetic text had not yet reached mainstream accessibility or commercial scale. While the Rules prohibited impersonation, misinformation, obscenity, and privacy violations, they did not specifically address the unique risks posed by AI-generated content that could realistically mimic real individuals or events. As deepfake tools became widely available and incidents of AI-powered financial fraud, non-consensual intimate imagery, political misinformation, and executive impersonation increased,[3] enforcement under the present framework became interpretive rather than explicit.

The 2026 amendment addresses this gap by formally introducing the concept of "Synthetically Generated Information" (**SGI**) under Rule 2(1)(wa), defined as audio, visual, or audio-visual information that is artificially or algorithmically created, generated, modified, or altered using a computer resource, in a manner that such information appears to be real, authentic, or true and depicts or portrays any individual or event in a manner that is, or is likely to be perceived as, indistinguishable

---

[1] Ministry of Electronics and Information Technology, Gazette Notification G.S.R. 120(E), dated 10 February 2026.

[2] "Centre Notifies IT Rules Amendment To Regulate AI-Generated Content; Platforms Have to Take Down Illegal Content Within 3 Hrs," LiveLaw, 10 February 2026.

[3] Observer Research Foundation, "Deepfakes and Financial Cybercrime: India's Multi-Layered Response," 15 January 2026, available at orfonline.org.

SKV LAW OFFICES

from a natural person or real-world event. The amendment represents more than a gap-filling exercise. It constitutes a conceptual shift in regulatory philosophy, from regulating what content is posted to regulating how content is created. Indian digital law is no longer only reactive to content harms; it is now anticipatory of technological capacities. Beyond defining SGI, the amendment strengthens transparency requirements and considerably reduces takedown timelines.

Synthetically Generated Information: A Perceptual, Not Technical, Threshold

This definition captures deepfakes, AI-generated voice clones, synthetic avatars, and other generative AI outputs capable of impersonation or deception. Two features of the definition deserve particular attention. First, the language "artificially or algorithmically created" is technologically neutral. It covers generative AI, neural networks, and rule-based automation without tying the definition to a particular method. Second, the threshold is perceptual rather than technical. The test is not whether a particular technology was used, but whether the output appears real and is likely to be perceived as indistinguishable from authentic content. This deception-focused threshold aligns with global deepfake definitions and has major implications: content may fall within the regulatory scope based on how it is perceived, even in the absence of proof of the creator's deceptive intent.

It is worth noting that the October 2025 draft applied to information that "reasonably appears to be authentic or true" without limiting the form of such content. The final notification confines the definition to audio, visual, or audio-visual material, a narrowing that reflects stakeholder responses.[4] Recognizing industry concerns regarding overbreadth, the amendment carves out three categories of activity from the SGI definition: routine editing activities such as formatting, color adjustment, noise reduction, transcription, or compression that do not materially alter the substance, context, or meaning of the content; routine creation of documents, presentations, educational materials, and research outputs using illustrative or template-based content, provided no false document is created; and the use of computer resources solely to improve accessibility, clarity, quality, translation, or searchability without manipulating the underlying content.[5] Each carve-out uses the language of good faith and absence of material distortion, a deliberate effort to draw a line between deceptive simulation and functional digitization.

When Synthetic Becomes Unlawful

It is important to emphasize that AI-created content is not unlawful merely because it is synthetic. It becomes unlawful when it violates

---

[4] "MeitY Notifies New Amendments to IT Rules on Synthetic Media," Medianama, 10 February 2026.

[5] MeitY, Frequently Asked Questions on the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, available at meity.gov.in.

applicable law or falls within prohibited categories under the Rules. The term "unlawful content" is defined by the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, read together with the IT Act and other applicable Indian laws.

Under Rule 3, intermediaries must not host or transmit content that is, among other things, defamatory, obscene, sexually explicit, invasive of privacy, threatening to national security or public order, or that impersonates another person. Violations under the IT Act and other applicable laws, including criminal law, copyright law, and election law, may independently render content unlawful. References to the Indian Penal Code have now been replaced with the Bharatiya Nyaya Sanhita, 2023, aligning the framework with India's updated criminal law architecture.

The amendment goes further by recognizing specific categories of prohibited SGI under Rule 3(3). Intermediaries offering computer resources that enable the creation or dissemination of SGI must deploy reasonable and appropriate technical measures, including automated tools, to prevent the generation of SGI that constitutes child sexual exploitative and abuse material, non-consensual intimate imagery, or obscene and sexually explicit content; content creating false documents or false electronic records; content relating to explosives, arms, or ammunition; and content falsely depicting individuals or events in a manner likely to deceive. For example, a clearly labelled artistic AI avatar would not ordinarily fall within these prohibitions, whereas a deepfake impersonating a corporate executive for financial fraud would likely trigger regulatory consequences.

Hosts and Enablers: A Two-Tier Compliance Structure

The amendment imposes a two-tier compliance framework depending on whether an intermediary merely hosts SGI or actively enables its creation, generation, modification, alteration, publication, transmission, sharing, or dissemination. This distinction is significant because the compliance burden differs materially depending on which side of the line a platform falls.

Intermediaries that merely host user-generated content continue to operate under the general due diligence framework, with compressed takedown timelines and expanded notification obligations. Intermediaries that actively enable the creation of synthetic content face a substantially heavier set of obligations, including the deployment of automated detection tools, active prevention of prohibited SGI, mandatory labelling and metadata embedding, and additional user notification requirements under Rule 3(1)(ca). These additional notices must warn users that violations may attract penalties under the Representation of the People Act, 1951, the Indecent Representation of Women Act, 1986, the Sexual Harassment at Workplace Act, 2013, and the Immoral Traffic Act, 1956, and can lead to immediate disabling of access, account suspension, identity disclosure to victims, and reporting to appropriate authorities.

This targeting of the tools of creation, rather than the content produced, constitutes a regulatory approach similar to the governance of dual-

use technologies, in which the capability to cause harm, rather than the harm itself, attracts regulatory attention.

Transparency, Not Prohibition: The Labelling Regime

The amendment creates a transparency obligation requiring intermediaries to ensure that synthetic content that does not fall within the prohibited categories is clearly and prominently labelled. Platforms must implement visible disclosures for visual content and audio disclosures for audio content. Where technically feasible, intermediaries must embed permanent metadata and unique identifiers to trace the computer resource used to create, generate, modify, or alter the SGI. Critically, intermediaries cannot remove these labels or metadata.

The objective is to enable users to distinguish synthetic media from authentic content, thus lowering the risk of deception. This creates a transparency regime rather than a ban on synthetic speech, in line with global approaches such as the EU's provenance proposals and C2PA standards.[6]

Pre-Publication Verification: The SSMI Burden

Under new Rule 4(1A), Significant Social Media Intermediaries (SSMIs) that enable displaying, uploading, or publishing information on their platforms face additional obligations. SSMIs must require users to declare whether content is SGI before publication, deploy automated tools to verify such declarations, and ensure that verified SGI is clearly labelled with appropriate notices.

This pre-publication declaration mechanism is one of the more operationally significant features of the amendment. It places verification responsibilities on platforms at the point of upload rather than after dissemination. However, the verification standard remains ambiguous. The Rules frame these duties as requiring "reasonable and appropriate" measures, but no performance benchmarks, acceptable error-rate thresholds, or specific technical standards have been prescribed. This interpretive gap is likely to give rise to compliance ambiguity and possible litigation as the framework matures.

The Three-Hour Takedown and Tiered Response Architecture

One of the most significant changes introduced by the amendment is the reduction of takedown and grievance timelines across multiple categories. Under the earlier framework, intermediaries were generally required to remove unlawful content within 36 hours of receiving actual knowledge or a valid notice.

The amended Rules now impose a tiered structure. For the most

[6] European Commission, Draft Code of Practice on Transparency of AI-Generated Content, 17 December 2025. See also Natalia Garina, "What the EU's New AI Code of Practice Means for Labeling Deepfakes," TechPolicy.Press, 7 January 2026.

sensitive category, content exposing private areas, nudity, sexual acts, or artificially morphed images, intermediaries must act within two hours (reduced from 24 hours). For other unlawful content, including unlawful SGI, the timeline is three hours from receipt of a valid order from the Government or a court (reduced from 36 hours). Complaints relating to intimate images or content of an individual require action within 36 hours (reduced from 72 hours). General user grievances must now be resolved within 7 days (down from 15).

The amendment also tightens the authorization procedures for takedown orders. Notices must be issued "by order in writing," and for police administration, authorized officers must be at least of Deputy Inspector General rank. This provides a procedural safeguard against overbroad or informal takedown demands, though the compressed timelines will, in practice, require intermediaries to maintain round-the-clock monitoring, clearly defined escalation procedures, and technical systems capable of quick action.

The Unsettled Questions: Ambiguity in Application

Despite incorporating stakeholder feedback following the public consultation on draft rules released in October 2025, the amendment retains certain interpretive challenges. Several notable changes were made between the draft and the final notification. The definition of SGI was narrowed from any information that "reasonably appears to be authentic or true" to audio, visual, or audio-visual material only, effectively excluding text-only AI outputs. The draft's proposed minimum labelling threshold of 10 per cent of visual surface area or audio duration was removed, leaving platforms to determine display standards. The compressed takedown timelines of two and three hours were not present in the October 2025 draft, which had retained the existing 36-hour window. Safe harbor protections were also broadened, with the final version explicitly clarifying that proactive moderation through automated tools does not, by itself, jeopardize immunity under Section 79.

However, certain gaps persist. The standard of "reasonable and appropriate technical measures" for detecting prohibited SGI remains undefined, with no performance benchmarks or acceptable error-rate thresholds. The mechanism for authenticating user declarations required of SSMIs is similarly unclear. The carve-outs for "routine editing" and "good-faith creation" remain subject to interpretation, particularly in relation to satire, parody, or artistic expression.

These ambiguities are not purely academic. Platforms will be required to make real-time classification decisions under compressed timelines, and the consequences of error, whether via over-removal or under-enforcement, carry considerable legal exposure. Civil society organizations have already argued that the compressed timelines eliminate meaningful human review, effectively forcing platforms toward automated over-removal.[7] Close engagement with regulators will be essential as the framework matures.

SKV LAW OFFICES

---

[7] Internet Freedom Foundation, "IT Intermediary Amendment Rules, 2026 contradict their purpose," 12 February 2026, available at internetfreedom.in.

Cross-Sector Implications and the Proportionality Gap

While social media platforms are most directly affected, the amendment has cross-sector implications. Generative AI tool providers, digital publishers, OTT platforms, financial institutions, influencer marketing agencies, gaming platforms, and cloud hosting providers should all assess their exposure to obligations related to deepfakes.[8] Any organization that enables user-generated content, deploys AI to create realistic representations of individuals, or relies heavily on digital communication at scale should undertake a quick review of its AI governance and content moderation systems. For smaller intermediaries and startups, the compliance burden may be disproportionate, given that the same obligations apply irrespective of platform size, and the ten-day implementation window between notification and enforcement leaves minimal room for transition.[9]

Constitutional Tensions: Speech, Privacy, and Innovation

The amendment raises several questions at the intersection of regulatory policy and constitutional principle. Mandatory labelling of synthetic content, while framed as a disclosure and consumer protection measure, may experience challenges as compelled speech under Article 19(1)(a). Courts may, however, uphold it as a reasonable restriction in the interest of preventing deception and defending public order. The metadata embedding and traceability requirements raise privacy concerns, particularly where provenance tracking could be used to identify creators of lawful content. The compliance costs associated with automated detection, labelling infrastructure, and compressed timelines may dampen AI innovation, particularly for early-stage companies. Without differentiation based on platform size or maturity, the framework risks imposing institutional-grade obligations on nascent enterprises. These tensions will ultimately be tested through judicial interpretation and regulatory practice.

India in Global Context: Enforcement Over Licensing

Globally, jurisdictions are increasingly regulating synthetic media and AI-created content. The European Union has adopted a risk-based governance framework under the EU AI Act, with transparency obligations for AI-generated content set to be fully enforced in August 2026. The EU's recently published draft Code of Practice on Transparency of AI-Generated Content provides practical guidance on labelling, watermarking, metadata, and disclosure measures, serving as a link between self-regulation and binding rules under the Digital Services Act. The United States continues to rely on a disjointed approach, with state-level deepfake laws and enforcement through consumer protection authorities, without fixed federal takedown

---

[8] Ministry of Electronics and Information Technology, Gazette Notification G.S.R. 120(E), dated 10 February 2026.

[9] "Centre Notifies IT Rules Amendment To Regulate AI-Generated Content; Platforms Have to Take Down Illegal Content Within 3 Hrs," LiveLaw, 10 February 2026.

timelines. China has introduced strict labelling and identity verification requirements for generative AI services within a more state-controlled regulatory model.

India's approach is intermediary-focused and enforcement-driven. While it does not introduce a comprehensive AI licensing regime, it imposes one of the shortest mandatory removal timelines globally through the tiered takedown structure and links compliance directly to safe harbor protection. Unlike the EU, which differentiates obligations based on platform scale through its "very large online platform" category, India's framework applies uniformly, creating operational challenges for smaller intermediaries. India, therefore, functions as a high-speed enforcement jurisdiction within the developing global scene of synthetic media regulation.

The Era of Synthetic Reality

The 2026 amendment represents a structural shift in India's digital regulatory framework. By defining synthetic content, clarifying the scope of unlawful content, mandating transparency through labelling and provenance tracking, imposing tiered takedown requirements, and distinguishing between platforms that host and platforms that enable synthetic media, the Government has significantly heightened intermediary accountability. Intermediaries are no longer passive conduits; they are governance actors within the digital ecosystem.

Businesses operating in India should treat this development as a priority compliance matter and conduct a structured assessment of their AI deployment practices, content moderation frameworks, and rapid-response capabilities. The success of this system will ultimately depend on effective implementation, technological feasibility, and judicial interpretation, but it is clear that Indian digital law has now entered the era of synthetic reality.