Balancing Urgency and Liberty: Constitutional Scrutiny of India's emerging blueprint for regulating AI-generated content

Authored by Sanskriti Shrimali (Associate at Kochhar & Co.) and co-authored by Samiron Borkataky (Partner at Kochhar & Co.)

Published in LiveLaw

As New Delhi embraced the cascading effects of the India-AI Summit 2026 and the new amendment of 10.02.2026 to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021[1] ("**2026 Amendment**")[2] in the same month, its ripple effect will be seen in balancing the rights to digital privacy with commercial objectives. With the recent influx of petitions to safeguard the personality rights of celebrities against commercial exploitation through AI-generated depictions or to regulate satirical content amplification against present standards of morality, the debate about Privacy and Policy not aligning with constitutional jurisprudence in India's AI Transition has been reignited.

While Indian law and policy framework is playing catch-up with the fast-evolving media spaces led by AI generative content, the legal contours of personality rights *vis-à-vis* commercial innovation and free speech concerns in India, continue to be at a nebulous stage. Up until last year, the country's data-driven market was oscillating between the initial Information Technology regime and the forthcoming Digital Personal Data Protection Act("**DPDP Act**") and has now suddenly seen a relatively swift introduction of the 2026 Amendment. Whilst at first brush, this may land as a much-needed welcome step, but it appears less as the culmination of a fully stabilised digital protection ecosystem and more as a reactive intervention. The regulatory thresholds have left intermediaries with an unsettled and evolving compliance architecture.

The Revised Take-down obligation

In the hype to protect one's digital personalities, particularly against deepfakes, the 2026 Amendment has now lowered the threshold for take-down of content, specifically in relation to "*synthetically generated information*" (SGI) i.e., AI generated art, videos, audio, etc, unless they conform to the "good faith" standard.[3] However, the onus to interpret this broad term with the said standard is *inter alia* left open to the intermediaries (*YouTube, Meta, etc.*)

[1]Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E) (India).
[2]Ministry of Electronics & Information Technology, Notification, G.S.R. 120(E), Gazette of India Extraordinary Part II, New Delhi, Feb. 10, 2026 (India).
[3]Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, r. 2(1)(wa), G.S.R. 120(E) (India).

This new amendment has now compressed the timeline to takedown potentially harmful content as per Rule 3(b) and (d),[4] to merely 3 hours from 72 hours i.e., a 92% reduction in response time after receiving notice either *via* court order or the relevant authority. While one could argue that this may incentivise intermediaries to pre-emptively remove potentially harmful content and undertake deliberate review, this also has the potential of making the Intermediaries "*passive conduits*". Furthermore, the safe harbour protection enjoyed by the Intermediaries under Section 79 of the IT Act, 2000, will now be subject to their active response to the revised take-down and grievance timeline, compliance with the added obligations and good faith. The grey area here however, is that the amendment does not explain what happens if these commercially driven intermediaries fail to meet the deadlines or the disputes are mishandled due to technical errors. Such omissions/uncertainties will risk the safe harbour protection and might make intermediaries more conservative to avoid any legal scrutiny.

In addition, earlier, the IT Rules 2021 had used terms like "*endeavour to deploy*" technology-based measures to verify the accuracy of the declaration by the users, which makes it directory.[5] However, the current 2026 Amendment uses the term "*shall*", making it mandatory. This can be applauded as a forward-looking measure and creates a serious approach to handling AI-led misinformation. In furtherance of the said requirements, platforms now must incorporate technical measures to accurately verify and clearly label the AI-generated content before it goes live. However, when it comes to determining the standard of accuracy of a user's declaration on synthetically generated content under Rule 4(1A) of the 2026 Amendment, the same is again not clearly defined.[6] The Intermediaries are yet to figure out the parameters on which such accuracy will be checked and what would be the threshold to demonstrate compliance by the creators.

While the above looks attractive, does it meet the vires test?

Strictly analysing from a constitutional perspective, this imposition of quick redressal may push the regulatory framework towards prior restraint, for blocking a content even before publication, which may impinge upon a citizen's fundamental right to freedom of speech and expression, as it will be subject to clearance by private intermediaries before entering the public domain. It is understandable that the objective is prevention at source, but merely flagging the synthetic origin (for the sake of compliance) without any assessment on illegality or misinformation may not serve the ends of justice. There is a lack of guidance in the assessment of such content, which might delay or discourage dialogue rather than addressing the consequences after publication of alleged misleading content.

Furthermore, the acting mechanism against the AI-generated content is a quick takedown, but it being the only method to curb the harm caused by misleading SGI is far-fetched. It rather appears to create an imbalance of rights and responsibilities and would have a direct impact on the right to freedom of speech. The Supreme Court's decision in *Shreya Singhal v. Union of India* remains instructive in the present context. While striking down Section 66A of the IT Act 2000, for vagueness and overbreadth, the Court highlighted that restrictions on online speech must fall strictly within the scope of Article 19(2).

The intermediaries will now be incentivised to censor without much reasoning to meet the newly set deadlines, which were never part of the initial draft of the amendment proposed by the government in 2025.[7] It is also interesting to note that while promoting the "*Democratic diffusion of AI*" at the Impact Submit 2026, the government did not constructively consult the

[4]Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, r. 3(1)(b)(ii), 3(1)(d), G.S.R. 120(E) (India).

[5] Ministry of Electronics & Information Technology, Explanatory Note: Proposed Amendments to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 in Relation to Synthetically Generated Information, Oct. 22, 2025.

[6]Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, r. 4(1A), G.S.R. 120(E) (India).

[7]Ministry of Electronics & Information Technology, Draft Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2025, Oct. 2025.

stakeholders on the additional amendments, which went beyond the ones proposed before. This may have resulted in the insertion of the new sub-clauses to Rule 3 that impose due diligence obligations in relation to SGI, particularly where content either depicts a person or an event "*in a manner that is likely to deceive*".[8]

This imbalance of rights by skewing power to block accounts and take down content is further exacerbated by the central government's recent notification on 30.03.2026, calling for public consultation on additional amendments to the IT Rules 2021, as these would confer explicit power to the executive/ministry over online content regulation.

Last year, the Supreme Court took to the whip to put in order the issue of algorithmic amplification of content creation and its impact on free speech. *First*, in March 2025, it stayed the Union Government's notification that established a fact check unit under Rule 3(1)(b)(v) of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2023 for the usage of vague terms like "*Fake"* and *"Misleading*", which was resulting in unchecked discretion on the fact-checking units before elections. *Second*, in August 2025, the Court expressly advised the Union Government that any framework to regulate online content should not be reactionary, but responsive. In other words, laws should not curtail the freedom of speech and expression as a mere response to a singular incident, such as in the case of "*India's Got Latent*". In the contemporary setup, comedy and satire provide a "*safety valve*" for sensitive social issues and have emerged as a fifth pillar of a healthy democracy.

In this constantly evolving terrain of technological innovation, commercial incentives and administrative restraint, the questions of proportionality, procedural fairness, and narrowly tailored enforcement remain central to any assessment under Articles 19(1)(a) and Article 21 of the Constitution. However, in the present scheme of things, the State will be the deciding authority as to what remains online, and any contrary perspective will take a back seat. While the creator economy has experienced a rapid surge, the 2026 Amendment places the reins firmly in the hands of the State, potentially curbing creativity, expression, and the dynamic exchange of ideas that fuels innovation. In effect, the executive authority becomes the primary arbiter of online legality, and intermediaries, facing liability risks, are incentivised to comply mechanically, thereby recreating the very chilling effect that the Court cautioned against. Further, it appears that in its zeal to curb the problem of deep fakes and misinformation, the government has somehow also missed the bus, in so much as the 2026 Amendment clearly lacks the very threshold of content classification between *satire vs. misleading content vs. morality*, as discussed widely all over the country.

In view of the overbreadth and enforcement challenges of the 2026 Amendment, the focus should shift from reactive takedown mechanisms to a more holistic framework. One such sustainable approach can be active mitigation by addressing the platform design/ recommendation algorithm of these intermediaries themselves by way of algorithmic audits, reducing how harmful content is "amplified and transmitted" based on the number of likes, comments and shares, in the first place. This can be done by complementing content regulation to the data governance framework, i.e., the DPDP Act. As data is the oil to keep the algorithmic activities up and running, therefore the use of the collected data should be restrictive and transparent. This will directly strike the precision of recommendation systems, which will eventually reduce the circumstances where users are being pushed into echo chambers or extremist content loops. While this cannot address the biggest challenge of identifying harmful content at source, it can potentially limit this reach and engagement. As the harmful content cannot be perfectly defined, it can be divided into categories: a) Explicitly illegal, such as child abuse, terrorism; and b) Legal, but can damage social values, such as misleading information, dark humour or hate speech.

[8]Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, r. 3(1)(iv), G.S.R. 120(E) (India).

Jurisdictions such as the European Union, through the Digital Services Act, address the latter by managing the amplification or systemic risk, and Australia, through initiatives led by the e-Safety Commission, by embedding safety considerations directly into platform architecture. Such architectural mitigation requires a deep understanding of the algorithms making any content viral and a significant in-house capacity building, for both implementation and compliance. The unresolved question, however, is whether sufficient alignment of incentives exists among the State, platforms, and creators to prioritise long-term systemic safety over short-term compliance and commercial gains.