

180 Minutes to Respond: India's New Regulatory Framework for Deepfakes and Synthetic Media

AUTHORED BY:

Mr. Vedant Saraf

PARTNER

CO-AUTHORED BY:

Ms. Megha Paul

ASSOCIATE



INTRODUCTION

Imagine waking up to several missed calls from friends, colleagues and family. You open your phone and discover that a video appearing to show you has already been circulating online for 2 (two) hours. Your face is recognisable. Your voice sounds like yours. But the words, conduct and actions depicted in the video are entirely fabricated. By the time you realise what has happened, thousands of people may already have seen it. That is the particular difficulty with deepfakes. The harm is not limited to the creation of false content. Once released online, synthetic media can travel faster than a victim can respond, and correcting a false impression may be considerably harder than creating it.

India's regulatory framework has now responded to this problem with a more specific set of obligations. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 ("**2026 Amendment Rules**"), notified on February 10, 2026 and brought into force on February 20, 2026, introduce a dedicated framework for "synthetically generated information" ("**SGI**"), alongside tighter timelines for content removal and grievance redressal.

The changes are significant because they move beyond treating harmful synthetic content simply as another category of unlawful online information. They introduce obligations aimed at prevention, labelling, provenance and verification, while also shortening the time within which certain content must be addressed.

WHAT HAS CHANGED

The 2026 Amendment Rules define SGI to cover audio, visual and audio-visual information that is artificially or algorithmically created, generated, modified or altered using a computer resource in a manner that appears real, authentic or true and is, or is likely to be perceived as, indistinguishable from a natural person or real-world event. The definition is therefore broader than the popular term "deepfake", although deepfakes would ordinarily fall within it where they meet the prescribed threshold.

The Rules also recognise that not every use of AI or computer-based editing should be treated as synthetic media requiring the same regulatory treatment. Routine or good-faith editing, enhancement, technical correction, colour adjustment, noise reduction, transcription and compression, among other specified uses, are excluded where they do not materially alter or misrepresent the underlying content.

This distinction matters. The regulatory concern is not the use of AI or digital editing as such, but the creation or alteration of content in circumstances where its apparent authenticity can mislead users or facilitate unlawful conduct.

THE THREE-HOUR CLOCK IS NOT A BLANKET DEEPFAKE TAKEDOWN RULE

One of the most visible changes is the reduction of the removal timeline under Rule 3(1)(d) from 36 (thirty-six) hours to 3 (three) hours.

Where an intermediary receives the prescribed "actual knowledge" of unlawful information, it must remove or disable access to the information within three hours. Importantly, actual knowledge for this purpose does not arise from every complaint received by a platform. The Rules prescribe the circumstances in which it arises, including an order of a court of competent jurisdiction or a reasoned written intimation issued by an appropriately authorised government officer and satisfying the specified requirements. The distinction is

important because the three-hour period should not be understood as a universal rule requiring every alleged deepfake to be removed within three hours of a user complaint. There is, however, a separate and particularly relevant two-hour mechanism.

Under Rule 3(2)(b), where an individual or a person acting on their behalf complains about content that is *prima facie* in the nature of material exposing the individual's private area, depicting full or partial nudity, depicting a sexual act or conduct, or involving electronic impersonation, including artificially morphed images, the intermediary must take all reasonable and practicable measures to remove or disable access within 2 (two) hours of receiving the complaint. For victims of non-consensual intimate imagery or certain forms of deepfake impersonation, this provision can therefore be particularly significant.

The 2026 amendments have also reduced other grievance-redressal timelines. The general resolution period under Rule 3(2)(a) has been reduced to 7 (seven) days, while the specified period for certain removal requests has been reduced from 72 (seventy-two) hours to 36 (thirty-six) hours.

THE NEW SGI FRAMEWORK GOES BEYOND TAKEDOWN

The more consequential change may lie in the new Rule 3(3), which imposes dedicated due-diligence obligations in relation to SGI. Where an intermediary provides a computer resource that enables or facilitates the creation, generation, modification, publication or dissemination of SGI, it must deploy reasonable and appropriate technical measures, including automated tools or other suitable mechanisms, to prevent users from creating or disseminating specified categories of unlawful SGI. These include, among other things, child sexual exploitative and abuse material, non-consensual intimate imagery, certain obscene or sexually explicit material, false documents or electronic records, content relating to the preparation or procurement of explosives, arms or ammunition, and synthetic content that falsely depicts a natural person or real-world event by misrepresenting the person's identity, voice, conduct, action or statement in a manner likely to deceive.

For SGI that does not fall within those prohibited categories, the regulatory approach is different. The content must be prominently labelled as synthetically generated, and, to the extent technically feasible, must contain permanent metadata or another appropriate technical provenance mechanism, including a unique identifier capable of identifying the intermediary's computer resource used to create, generate, modify or alter the content. The intermediary must also not enable the removal or suppression of the prescribed label or provenance information. This is an important shift in regulatory thinking. The framework is no longer concerned only with what happens after harmful content goes viral. It also seeks to build transparency and traceability into the creation and distribution of synthetic media.

ADDITIONAL OBLIGATIONS FOR SIGNIFICANT SOCIAL MEDIA INTERMEDIARIES

The Rules go a step further for Significant Social Media Intermediaries ("**SSMIs**"). Before information is displayed, uploaded or published, an SSMI must require users to declare whether the information is synthetically generated. It must then deploy appropriate technical measures, including automated tools or other suitable mechanisms, to verify the accuracy of that declaration having regard to the nature, format and source of the information. Where the declaration or technical verification establishes that the content is SGI, the SSMI must ensure that it is clearly and prominently displayed with an appropriate label or notice identifying it as synthetic.

The significance of this obligation lies in the move from a purely reactive model towards an element of pre-publication verification for synthetic content on large social media platforms. The Rules nevertheless recognise the practical limits of technology. The obligation to verify declarations requires “appropriate” technical measures having regard to the nature, format and source of the information; it does not amount to a statutory guarantee that every synthetic item can be identified with certainty.

WHAT DOES THIS MEAN FOR SAFE HARBOUR?

Section 79 of the Information Technology Act, 2000, provides a conditional exemption from liability for intermediaries in respect of third-party information, subject to the conditions prescribed in the Act. In particular, the intermediary must satisfy the statutory requirements, including observing the prescribed due diligence.

The 2026 amendments expressly clarify that removal or disabling of access to information, including SGI, in compliance with the Rules, including through reasonable and appropriate technical measures or automated tools, will not by itself amount to a violation of the conditions in Section 79(2)(a) or (b). This clarification is commercially relevant. Platforms have historically had to navigate the tension between taking down problematic content and maintaining the conditions for intermediary protection. The amended Rules make clear that compliance-driven intervention, including the use of automated tools contemplated by the Rules, will not itself defeat the relevant safe-harbour conditions. That should not, however, be read as creating an absolute immunity for intermediaries or as making every instance of non-compliance automatically equivalent to substantive liability for user-generated content. Section 79 remains a conditional statutory protection, and the consequences of losing that protection must be considered separately from the underlying liability for the impugned content.

THE USER STILL HAS AN IMPORTANT ROLE

Regulation can shorten response times and impose stronger obligations on platforms, but neither can completely eliminate the harm caused by synthetic media. Users can reduce their exposure by limiting the amount of personal material that is publicly available, particularly photographs, videos and voice recordings that can be used to impersonate them. Privacy settings on social media platforms can reduce unnecessary exposure, although they cannot guarantee that previously published material will not be copied or reused.

Basic account security also matters. Strong, unique passwords and two-factor authentication can reduce the risk of account compromise, which can otherwise provide bad actors with access to personal photographs, communications and other material that may be misused. Equally important is what happens before a potentially harmful video is forwarded. A video that appears convincing is not necessarily authentic. Users should consider its source, corroborate significant claims through reliable sources and avoid forwarding material where its authenticity or legality is uncertain.

For a victim, speed is equally important. Preserving the relevant URL, screenshots, account details and other evidence before requesting removal can help establish what was published and where. Depending on the nature of the content, the victim may also have remedies under other applicable laws in addition to the intermediary grievance mechanism.

A FASTER RESPONSE, BUT NOT A COMPLETE SOLUTION

The 2026 amendments represent a meaningful development in India's regulation of synthetic media. They introduce a defined concept of SGI, impose preventive obligations for specified unlawful synthetic content, require labelling and provenance mechanisms for other SGI, introduce additional verification obligations for SSIMs and materially shorten certain removal and grievance-redressal timelines. But regulation cannot eliminate the underlying problem.

A deepfake may take minutes to create and seconds to share. Its legal consequences may take considerably longer to resolve. The effectiveness of the new framework will therefore depend not only on the text of the Rules, but also on the ability of intermediaries to implement workable technical systems, the accessibility of grievance mechanisms, the speed of governmental and judicial intervention where required, and the willingness of users to verify before they amplify. The practical lesson is simple: when synthetic media makes it increasingly difficult to trust what we see, the law can provide a faster route to intervention, but responsibility for limiting its reach cannot rest with platforms alone.