Rashmika Mandanna’s deepfake: Regulate AI, don’t ban it

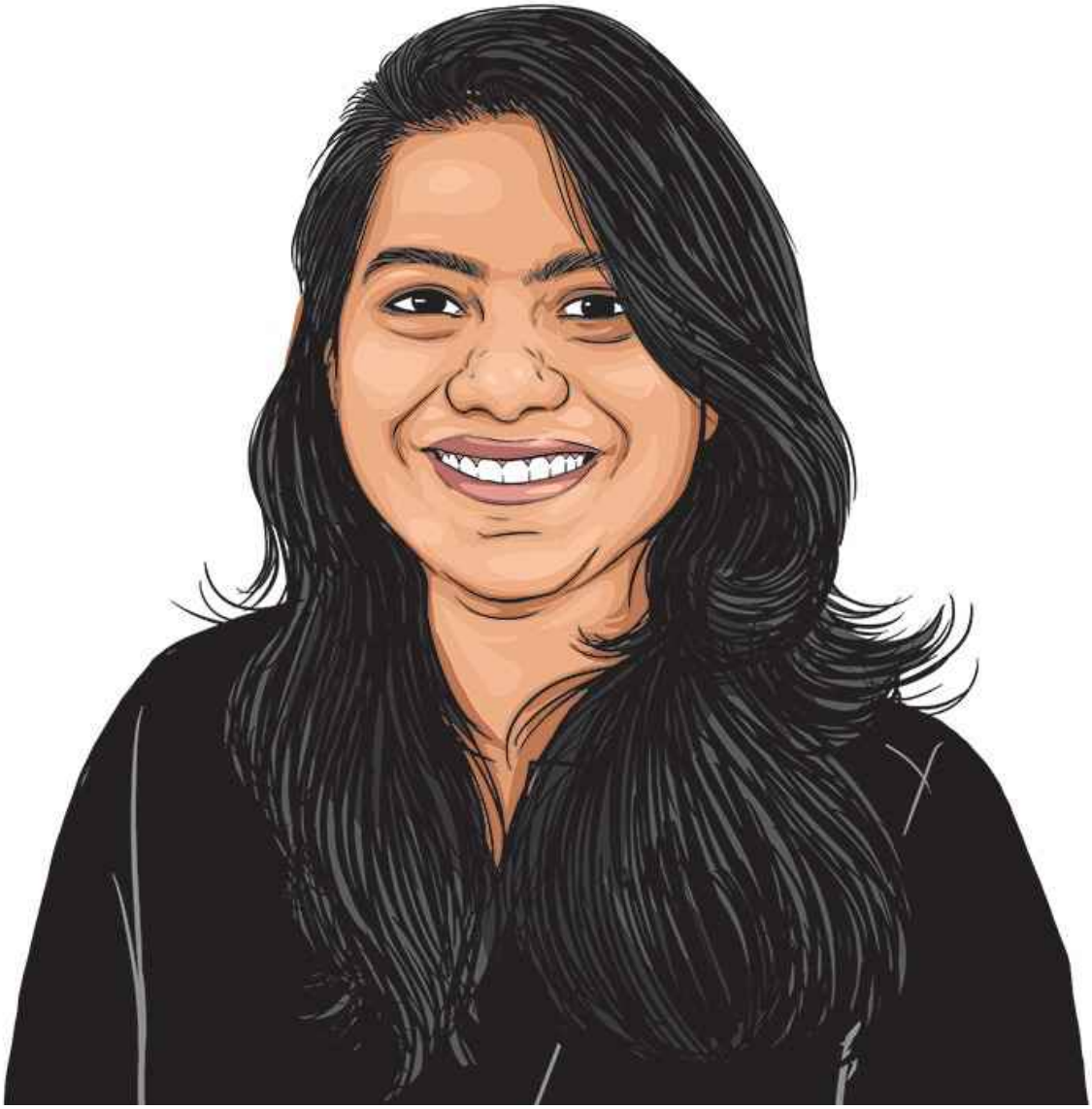
indianexpress.com/article/opinion/columns/rashmika-mandannas-deepfake-regulate-ai-dont-ban-it-9017666/

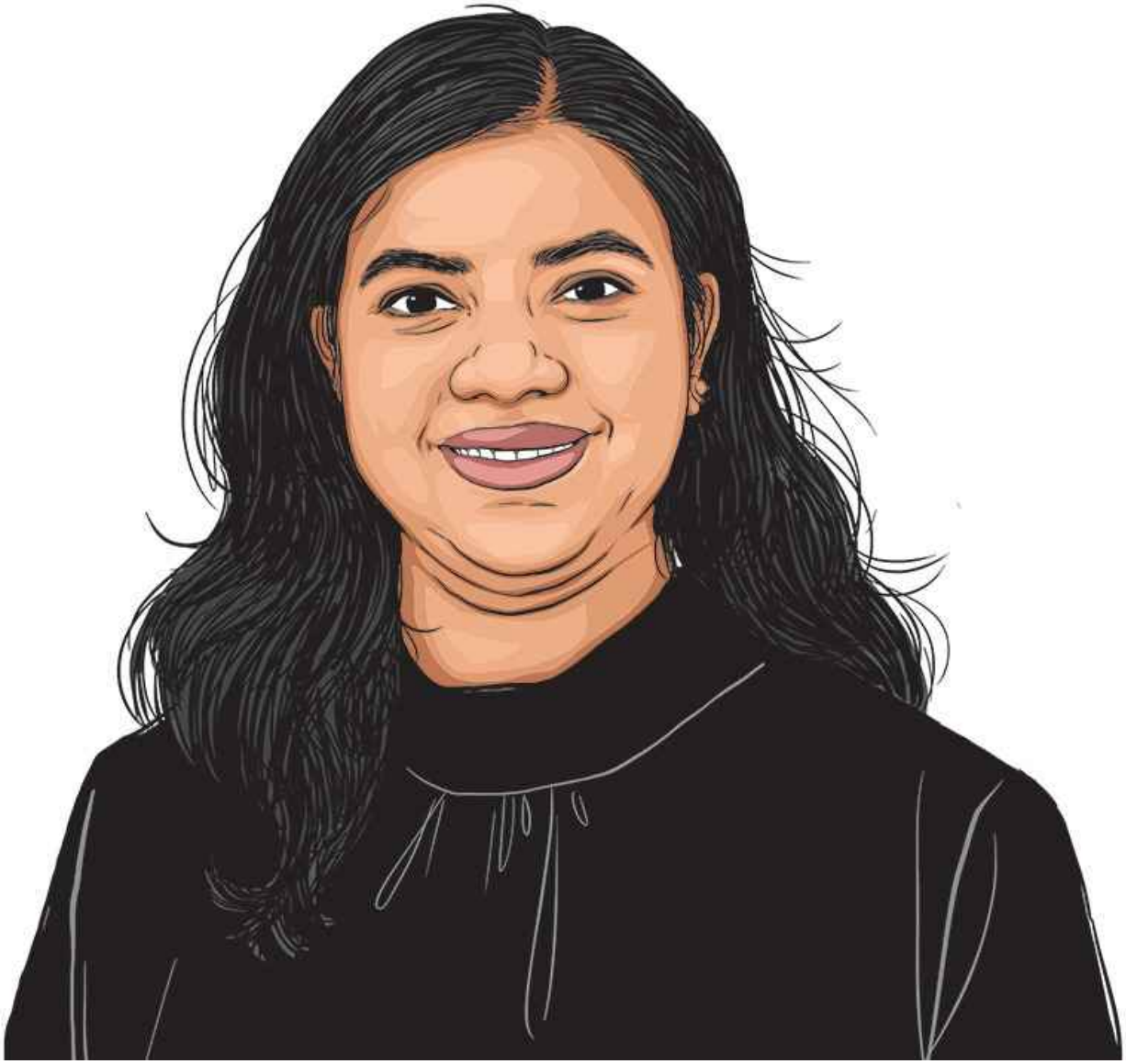
November 7, 2023

Opinion by Aishwarya Giridhar, Nidhi Singh

A holistic approach to the regulation of deepfakes should focus on the interplay between platform and AI regulation, and ways to incorporate safeguards for emerging technologies more broadly

Ultimately, the use of deepfakes echoes the issues we are likely to have with image-based generative AI and the scale and speed of how other kinds of problematic content can spread online. (Facebook/Rashmika Mandanna)





Aishwarya Giridhar

Nidhi Singh

Nov 8, 2023 09:21 IST

- Share

On November 5, fact-checker Alt News posted that a **viral video of actor Rashmika Mandanna** entering a lift was a deepfake. The video sparked much debate, with other actors calling for the legal regulation of deepfake videos. In response, Minister of State for Electronics and IT Rajeev Chandrasekhar talked about regulations under the IT Act,

which could tackle the spread of such videos. However, a holistic approach to the regulation of deepfakes should focus on the interplay between platform and AI regulation, and ways to incorporate safeguards for emerging technologies more broadly.

Deepfake content is created using advanced AI technology. While it may be used to generate fake videos, it can also be used to impersonate friends or loved ones to trick individuals into sending money to scammers. But there may also be legitimate uses for the underlying technology — for instance, to anonymise the voices and faces of journalists and help them remain safe in oppressive regimes. Therefore, a regulatory response that aims to draw a blanket ban on the use of such technology is likely to be disproportionate and possibly ineffective.

In Premium | Viral ‘Rashmika Mandanna video’ spotlights Big Tech’s deepfake problem, yet again

The life cycle of deepfakes can be divided into three parts – creation, dissemination and detection. AI regulation can be used to mitigate the creation of unlawful or non-consensual deepfakes. One of the ways in which countries such as China are approaching such regulation is to require providers of deepfake technologies to obtain consent of those in their videos, verify the identities of users, and offer recourse to them. The Canadian approach to prevent harm from deepfakes includes mass public awareness campaigns and possible legislation that would make creating and distributing deepfakes with malicious intent illegal. While there is no simple way to fix the problem, measures like adding watermarks to all AI-generated videos could be a good first step towards effective detection.

Detecting deepfake videos is becoming increasingly difficult. Due to advancements in AI content generation, the new generation of deepfakes is almost impossible to spot. This can significantly harm the people in the videos, and reduce trust in the credibility of video evidence. This is further complicated by the fact that creating deepfakes or false content in general is not by itself illegal, and may also be protected speech under the Constitution. Some content may be clearly unlawful — for example, identity theft or publishing content that violates intimate privacy on the internet. In other cases, illustrated by the context of the present video, it is unclear whether the video would be considered obscene, defamatory or merely a satirical impersonation. Consequently, many regulations around deepfakes focus on the sharing and dissemination of such content.

In India, the IT Act and related regulations address the content moderation obligations on online platforms. Although this framework would also apply to deepfakes, it is not entirely clear what actions platforms are required to take in this context.

Explained | ‘Deepfake’ video showing Rashmika Mandanna: How to identify fake videos

Typically, platforms must remove unlawful content within 36 hours of being notified by a court or government. If an individual is depicted in sexual acts or partial nudity or otherwise impersonated complains, platforms are required to remove such content within 24 hours. They are also required to publish terms of service that prohibit users from uploading content that impersonates other persons, and content that knowingly

communicates “misinformation”. Online platforms must also “make reasonable efforts to cause the user” to not upload such content on their platforms, and “act on” user complaints within 72 hours.

“Making reasonable efforts” to cause users to not upload such content is vague and ambiguous. Presumably, undertaking some kind of content moderation efforts would satisfy this requirement, meaning that most of the large social media platforms would be in compliance. Similarly, it’s not clear what “acting on” user complaints would mean, and whether actions such as down-ranking would suffice. This is important because the consequence of not complying with these rules could potentially make the platforms liable for such content.

Also Read | Rashmika Mandanna reacts to her viral deepfake video: ‘This is extremely scary…’

Ultimately, the use of deepfakes echoes the issues we are likely to have with image-based generative AI and the scale and speed of how other kinds of problematic content can spread online. It may be best to avoid reactionary calls for specialised regulation targeting the dissemination of deepfakes, and instead consider a multi-pronged regulatory response that engages with both AI and platform regulation. The upcoming Digital India Act, which will reportedly regulate AI and emerging technologies along with online platforms, provides an opportunity to address some of these issues.

Giridhar is project manager and Singh is programme officer, Centre for Communication Governance, National Law University, Delhi