THE THREAT OF DEEPFAKES

Relevant for: International Relations | Topic: India - USA

The protesters who created chaos in Capitol Hill on January 6 believe that the 2020 U.S. election was stolen by the Democrats. This is largely due to misinformation and disinformation of which deepfakes are a part. Deepfakes — synthetic media, meaning media (including images, audio and video) that are either manipulated or wholly generated by Artificial Intelligence — even have the power to threaten the electoral outcome of the world's oldest democracy. Sevreal social media platforms blocked President Donald Trump's accounts after the attack.

The cyberworld has been facing the challenge of deepfakes for a while now. AI is used for fabricating audios, videos and texts to show real people saying and doing things they never did, or creating new images and videos. These are done so convincingly that it is hard to detect what is fake and what is real. Detection can often be done only by AI-generated tools. Several books caution us against the threats of AI-generated content comprising non-existent personalities, synthetic datasets, unreal activities of real people, and content manipulation. Deepfakes can target anyone, anywhere. They are used to tarnish reputations, create mistrust, question facts, and spread propaganda.

In October 2020, the U.S. Senate summoned Facebook's Mark Zuckerberg, Twitter's Jack Dorsey and Google's Sundar Pichai to find out what they are doing to tackle online misinformation, disinformation and fabricated content. Senators said they were worried about both censorship and the spread of misinformation. According to Section 230 of the Communications Decency Act of 1996, a law which protects freedom of expression and innovation on the Internet, "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." This means that the companies are not responsible for the posts on their platforms. The chief executives said they need the law to moderate content, but industry watchers and some politicians feel that the law is outdated and needs to be revisited.

India also faces the same problem. So far, it has not enacted any specific legislation to deal with deepfakes, though there are some provisions in the Indian Penal Code that criminalise certain forms of online/social media content manipulation. The Information Technology Act, 2000 covers certain cybercrimes. But this law and the Information Technology Intermediary Guidelines (Amendment) Rules, 2018 are inadequate to deal with content manipulation on digital platforms. (The guidelines stipulate that due diligence must be observed by the intermediate companies for removal of illegal content.) In 2018, the government proposed rules to curtail the misuse of social networks. Social media companies voluntarily agreed to take action to prevent violations during the 2019 general election. The Election Commission issued instructions on social media use during election campaigns. But reports show that social media platforms like WhatsApp were used as "vehicles for misinformation and propaganda" by major political parties during the election.

This is worrying. Existing laws are clearly inadequate to safeguard individuals and entities against deepfakes. Only AI-generated tools can be effective in detection. As innovation in deepfakes gets better, AI-based automated tools must be invented accordingly. Blockchains are robust against many security threats and can be used to digitally sign and affirm the validity of a video or document. Educating media users about the capabilities of AI algorithms could help.

In July 2020, the University of Washington and Microsoft convened a workshop with experts to discuss how to prevent deepfake technology from adversely affecting the 2020 U.S. presidential

election. The workshop identified six themes: a) deepfakes must be contextualised within the broader framework of malicious manipulated media, computational propaganda and disinformation campaigns; b) deepfakes cause multidimensional issues which require a collaborative, multi-stakeholder response that require experts in every sector to find solutions; c) detecting deepfakes is hard; d) journalists need tools to scrutinise images, video and audio recordings for which they need training and resources; e) policymakers must understand how deepfakes can threaten polity, society, economy, culture, individuals and communities; and f) the idea that the mere existence of deepfakes causes enough distrust that any true evidence can be dismissed as fake is a major concern that needs to be addressed. In today's world, disinformation comes in varied forms, so no single technology can resolve the problem. As deepfakes evolve, AI-backed technological tools to detect and prevent them must also evolve.

K. P. Shashidharan is a former director-general in the office of the Comptroller and Auditor General of India

The Hindu has always stood for journalism that is in the public interest. At this difficult time, it becomes even more important that we have access to information that has a bearing on our health and well-being, our lives, and livelihoods. As a subscriber, you are not only a beneficiary of our work but also its enabler.

We also reiterate here the promise that our team of reporters, copy editors, fact-checkers, designers, and photographers will deliver quality journalism that stays away from vested interest and political propaganda.

Suresh Nambath

Please enter a valid email address.

To reassure Indian Muslims, the PM needs to state that the govt. will not conduct an exercise like NRC

You can support quality journalism by turning off ad blocker or purchase a subscription for unlimited access to The Hindu.

Sign up for a 30 day free trial.

END