The legal net to trap peddlers of deepfakes

Deepfakes rely on stealing and misusing a person's identity SHUTTERSTOCK

N S Nappinai

The most recent victim of sexually explicit Artificial Intelligence (AI) enabled deepfake videos being made viral is actor Alia Bhatt. Beware if you search to read more about this news, for an online search is likely to land you on porn websites catering to this category of fantasy, which is deepfake porn of celebrities.

From harming reputations through fake imagery, including nudity or sexually explicit acts, spreading false information or fake news, manipulating elections or public opinion about elected representatives or political figures, to committing financial frauds, deepfakes or misuse of AI have played a role. Law and technology would have to play a key role in preventing and protecting against these harms.

Often law is accused of not keeping pace with technological evolutions. This may not reflect the truth. The law is dynamic and applies, subject only to the limitations placed by the wording of a provision, to existing and evolving threats. The Indian Penal Code (IPC), 1861, which remains relevant even today is illustrative. Using deepfakes for criminal purposes may be an evolving threat, but using manipulated imagery including a modest photograph for harming a person, business or a nation, is not a new phenomenon. There is, therefore, no cause to believe that we do not have laws to combat this growing menace.

The Information Technology Act, 2000, as amended, (IT Act) and IPC provide substantial remedies against the abuse of deepfake technology. Sections 67 and 67A of the IT Act may be invoked for prosecuting the circulation of sexually explicit content of adults and Section 67B IT Act along with the POCSO Act for protecting children.

Deepfakes rely on stealing and misusing a person's identity or "unique identification feature", which can be penalised under Section 66C of the IT Act. Similarly, Section 66D of the IT Act may be invoked for financial fraud. Orders of the Delhi high court in the cases concerning Amitabh Bachchan and Anil Kapoor restricting abuse of their personality rights among several such cases from across jurisdictions, further strengthen victim rights in deepfake prosecutions. Forgery provisions under Sections 463 to 471 of the IPC may also apply to deepfake crimes. The offence tagged as "outraging the modesty of women" (Section 509 IPC) may be invoked. The Representation of People Act, 1951 and the Codes of Conduct mandated by the Election Commission may be invoked to combat deepfakes intended to manipulate elections. Cheating provisions under Section 420 (IPC) can combat financial fraud.

Preventive or protective measures from a victim's perspective would primarily be preventing uploads or dissemination of deepfakes or expedited takedowns. Presently, Section 79 of the IT Act read with the rules framed under the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, as of 2023 ("intermediary guidelines"), provides remedies for platforms to ensure the prevention of dissemination of harmful content affecting individuals and for expedited takedowns of nude or sexually explicit content within 24 hours of a report to the grievance officer of the platform used for circulating such content. Appeal goes to grievance appellate committees set up by the government, apart from remedies through criminal prosecutions or civil suits.

Expediting investigations would be helped through provisions that mandate tracing the provenance of content, from its creation to dissemination. This is resorted to under China's law as well as others such as the US and the European Union (EU). The US's proposed federal Deepfake Accountability Bill mandates watermarking to trace origin. This federal Bill with the acronym "Protect Act", mandates consent for circulating manipulated content on porn sites.

The EU's proposed AI Act 2023 uses risk categorisation of "unacceptable", "high" or "low" risk to regulate AI — generative AI including deepfakes fall under the "high risk" category — mandating inter alia transparency declarations that content is AI generated or fake. South Korea penalises deepfake-enabled crimes that harm the public interest. The UK awaits protection through its comprehensive Online Safety Bill.

India is contemplating specific and stringent provisions to combat deepfake online harms including provenance declaration through watermarking. Provisions may emerge as part of the proposed Digital India Act. The more immediate possibility, however, is through amendments to the intermediary guidelines, which may, at best, protect through further diligence mandates for platforms and expedited takedowns. Lessons from the Supreme Court order in *Re: Prajwala Letter dated 18.2.2015 Violent Videos and Recommendations* may be applied to use AI to combat this crime by blocking identified content from further circulation and expediting takedowns through explicit buttons for reporting. Simple solutions for this and other online harms may pave the way for a victim-centric approach.

N S Nappinai, an advocate practising in the Supreme Court, is founder of Cyber Saathi Foundation, a non-profit organisation. The views expressed are personal