Real or Fake? Dealing with Deepfakes Dilemma in Digital Society

Thought Leadership • February 4, 2025

'First published on Lexology'

The India Cyber Threat Report 2025 by the Data Security Council of India identified Deepfake exploitation as one of the critical cybersecurity threats faced in 2024[1]. Predictions for 2025 suggest Deepfakes will be extensively deployed in enhanced deception techniques, spreading malware and enabling highly sophisticated phishing attacks[2]. According to a 2023 McAfee study, 47% of Indians – highest globally – have either been victims or know someone affected by deepfake voice scams[3]. Issues of Deepfakes have targeted individuals across various domains, including celebrities[4], religious heads[5], and government leaders[6]. As such, this article undertakes a techno-legal analysis of Deepfake technology, aiming to explore effective measures for addressing the challenges posed by this burgeoning threat.

What are Deepfakes? – Technology, Development and Types

The EU AI Act 2024 defines Deepfakes as "AI-generated or manipulated image, audio, or video content resembling existing persons, objects, places, entities, or events, falsely appearing authentic."[7] In essence, Deepfakes are synthetic forgeries of media – videos, audios, images, or combinations – created using AI's deep learning techniques[8], designed to convincingly deceive viewers with their apparent authenticity.

Technologies for altering media have been evolving since the late 1990s. For instance, the 1997 Video Rewrite Program reanimated facial movements to match audio during film dubbing[9]. However, Deepfakes represents a significant leap in realism due to advancements in computer graphics and artificial intelligence. Introduced in 2014, Generative Adversarial Networks (GANs) remain the most common method for creating Deepfakes[10]. GANs involve two networks: a generator that creates artificial media and a discriminator trained on real data to classify the output as real or fake[11]. Iterative feedback refines the generator's creations, producing media indistinguishable from authentic content, even for sophisticated detection tools. Another technique, though less common, is Variational Autoencoders, where an encoder creates dense representations of input data, and a decoder recreates or manipulates the data[12]. This approach has been used for "face-swap" videos, replacing one person's face with another's in a video.

Deepfake technology became widely recognized in 2017 when a Reddit user named 'deepfake' released explicit fake videos of actresses, marking the beginning of its rapid evolution[13]. Today, Deepfakes can manipulate every human aspect in video – facial features, voice, and body

movements – even in real-time. Accessible to anyone, cheap tools and open-source projects have led to an explosion of hyper-realistic Deepfakes across social media, posing critical legal and ethical challenges.

Various kinds of Deepfakes exist today, targeting specific media elements or creating entirely synthetic content. Categories include face-swaps, lip-syncing, puppet-master manipulation, face synthesis and attribute manipulation, and audio Deepfakes[14]. Sophisticated tools like ZAO, FaceApp, ResembleAI etc. enable these techniques, raising concerns about their misuse and impact.

The good, The bad and The Ugly

Deepfakes technology, like any other innovation, offers a mix of opportunities and challenges. On the positive side, it significantly enhances educational experiences by vividly bringing historical figures and concepts to life, enriching student understanding[15]. It has become indispensable in the entertainment industry, particularly in computer-generated imagery, visual and special effects, driving creative storytelling to new heights[16]. Deepfakes also enable anonymous expression of opinions for journalists and citizens, without fear of losing their anonymity. Campaigns like David Beckham's malaria awareness video showcase its utility in delivering impactful messages across multiple languages[17]. Moreover, deepfakes play a vital role in medical therapy, assisting bereavement counselling[18] and aiding patients with speech impairments through voice restoration technologies.[19]

Conversely, the misuse of deepfakes poses grave threats to societal harmony and individual dignity. Deepfakes are widely exploited for creating Non-Consensual Intimate Images (NCII) and Child Sexual Abuse Exploitative Material (CSAEM)[20], constituting severe social evils. They are instrumental in corporate fraud, blackmail, and extortion, particularly targeting women with explicit content. Cybercriminals use deepfakes to spread misinformation, disrupt governance, and jeopardize national security. Fake evidence and altered media delay justice, while privacy violations erode personal autonomy. Even seemingly harmless creations, like satirical videos, can escalate into serious political or legal issues. Deepfakes amplify risks to commercial and intellectual property rights by enabling the creation of indistinguishable counterfeit goods that infringe trademarks, damaging brand reputation and increasing consumer confusion. These AI-generated counterfeits not only threaten brand integrity but also endanger public safety by bypassing quality controls, posing challenges for enforcement and legal adjudication[21]. Without proper regulation, deepfakes remain a perilous double-edged sword.

Efforts to curb issues of Deepfakes – Global Outlook

Many jurisdictions worldwide are taking the issue of deepfakes seriously, adopting varied legal approaches to address the challenges they pose and to find effective solutions. Analysing recent developments in jurisdictions worldwide reveals that nations are primarily addressing the mischief of

deepfakes through specific regulations targeting their misuse, rather than implementing a comprehensive governance model.

Jurisdictions like Australia[22], France[23], and South Korea[24] have amended their criminal laws to specifically criminalize the distribution of explicit images created using deepfakes. Canada is also actively deliberating changes to its criminal laws to address the challenges posed by deepfakes[25]. Australia's eSafety Commissioner and courts are taking stringent actions against offenders spreading deepfakes, even if the offenders are not residents of Australia[26]. Canadian[27] and South Korean[28] courts are addressing cases involving CSAEM deepfakes with urgency, delivering swift justice and imposing punishments on the higher end of the legal spectrum, reflecting a firm commitment to combating this serious issue and protecting vulnerable individuals.

China is addressing the challenges of deepfakes through stringent policy measures. In 2019, it introduced strict disclosure requirements, mandating individuals and organizations to clearly indicate the use of deepfakes in videos[29]. The law also prohibits distributing such videos without proper disclaimers. In 2023, additional provisions were enacted, requiring parties to obtain consent, verify identities, maintain government-registered records, report illegal content, and provide recourse mechanisms[30]. These provisions also mandate watermark disclaimers to ensure transparency. Notably, the Beijing Internet Court fined a face swap mobile application for violating individuals' personal rights[31], demonstrating China's steps to regulating deepfakes and protecting privacy rights.

The European Union addresses deepfakes under the EU AI Act, categorizing them as limited risk but imposing disclosure obligations on deployers sharing deepfake-generated content[32]. Additionally, the General Data Protection Regulation (GDPR) empowers data subjects to safeguard their rights against deepfake misuse, depending on the nature of the information involved. The Polish Data Protection Authority recently directed Meta to remove deepfake-generated advertisements that misused a data subject's information without consent[33]. The United Kingdom on the other hand has introduced broader legislation aimed at combating the spread of NCII created using deepfake technology[34].

The United States at the moment lacks federal legislation specifically targeting deepfakes, though proposals aim to address threats posed by the technology, provide legal remedies to victims, and mandate reports on digital content forgery advancements. Several states have enacted laws tackling issues like fabricating deceptive videos to harm election candidates, using deepfakes involving minors in sexual conduct, and requiring disclosure of synthetic media in election campaigns.[35] Courts across the country are strictly interpreting existing laws, imposing significant penalties on offenders in cases involving NCII and CSAEM created with deepfake technology, signalling serious judicial attention to this issue.[36]

Courts worldwide are addressing the misuse of deepfakes with seriousness, leveraging existing laws

to protect victims based on the specific nature and consequences of such misuse.[37]

Indian Stance

Although India considered introducing specific legislation to address the issue of deepfakes in late 2023, such a law has yet to be enacted[38]. The issue of deepfake has been a topic of parliamentary discussion since 2019, with the government asserting the adequacy of current provisions under the IT Act, 2000, including Sections 66C and 66D, which penalize identity theft and cheating[39]. Steps to combat deepfakes include collaboration between MeitY, MHA, and law enforcement, awareness programs like Information Security Education and Awareness (ISEA), and targeted workshops for state authorities. Notably, the IT Rules, 2021, along with subsequent amendments, impose strict obligations on intermediaries to detect, flag, and remove harmful content like deepfakes.[40] MeitY's zero-tolerance policy has driven consultations with industry stakeholders via Digital India Dialogues and led to an advisory mandating compliance with Rule 3(1)(b) of the IT Rules.[41] This includes educating users about prohibited content, enforcing reporting mechanisms, and ensuring swift action against violations. Intermediaries failing to adhere to these rules face loss of safe harbour protections under Section 79 of the IT Act and may attract liability under applicable laws.[42] By engaging public consultations and issuing periodic advisories, the government demonstrates its commitment to an Open, Safe, Trusted, and Accountable Internet through attempts for addressing the evolving threats of misinformation and deepfakes. In addition to the above measures, the government is periodically issuing directions and advisories based on existing information technology laws [43] and recommends the labelling of deepfake content to ensure transparency[44].

While India lacks specific legislation addressing deepfakes, various existing laws can apply depending on the nature of misuse. The Bharatiya Nyaya Sanhita includes provisions that can be considered as a direct prohibition to deepfake misuse.[45] Additionally, the Information Technology Act, 2000, is applicable in certain cases of digital mischief involving deepfakes[46]. The Digital Personal Data Protection Act, 2023, may also be relevant, as deepfake creation often involves unauthorized use of personal data[47]. The Copyright Act, 1957, could apply in circumstances such as copyright-protected material is used in training AI models or generating deepfakes. Moreover, the Dark Patterns Guidelines, 2023, address scenarios where deepfakes mislead consumers into commercial decisions they would not ordinarily make.[48] Law enforcement in India is actively combating the misuse of deepfake technology, arresting offenders involved in spreading unauthorized deepfakes of celebrities[49], coercing victims with explicit deepfake images[50], and using the technology for fraudulent activities[51].

Courts in India are increasingly grappling with the multifaceted challenges posed by deepfake technology, addressing its misuse across various contexts. In one instance, the Bombay High Court expressed grave concerns over deepfake bots on Telegram that enabled users to create explicit

content.[52] While recognizing the severe implications of such tools, the Court emphasized the need for the central government to identify and block these technologies.

The Delhi High Court, on multiple occasions, has taken steps to restrain the dissemination of deepfake videos. In a notable case, it prohibited the circulation of deepfake videos falsely portraying a lawyer being assaulted for representing a client. The Court held that such videos not only harm the lawyer's reputation but also pose a persistent threat of future misuse, potentially causing irreparable loss and injury[53]. Similarly, the Court acted decisively in another case to protect a prominent businessman whose identity was being fraudulently used in deepfake videos to deceive the public into making fraudulent investments[54]. Courts have also extended protection to celebrities[55] whose personality rights and fictional characters[56] whose commercial rights were violated through deepfakes.

In criminal cases involving deepfakes, courts have recognized the potential for harm and have adopted a strict approach. For instance, bail has been denied in cases where deepfake videos were disseminated[57], reflecting the judiciary's acknowledgment of the gravity of such offenses. Additionally, courts have observed the necessity of proving the authenticity of documents in disputes[58], cautioning against the use of manipulated evidence created through deepfake technology.

The Uttarakhand High Court addressed a case where an accused threatened to disseminate a deepfake video of the informant to deter them from contesting elections. By refusing to quash the first information report (FIR), the Court affirmed the need for a thorough investigation, highlighting its resolve to ensure accountability for threats involving deepfakes[59]. In another instance, the Bombay High Court directed social media platforms to remove deepfake videos targeting the managing director and CEO of the National Stock Exchange of India. These videos, which mimicked the plaintiff's voice and expressions to solicit public participation in fraudulent schemes, were ordered to be promptly taken down.[60]

Political misuse of deepfakes has also drawn judicial attention. Ahead of the 2024 general elections, the Delhi High Court opined that the Election Commission of India is the appropriate authority to address concerns about targeted deepfakes in the political arena[61]. Simultaneously, the Court called for governmental responses to public interest litigations seeking stricter regulations against websites facilitating deepfake misuse. Based on such matters, the central ministry has undertaken various initiatives to explore technical solutions to curb issues arising from deepfakes and has constituted committees to propose comprehensive measures to address these challenges.[62]

Overall, Indian courts are actively employing existing legal frameworks to address the evolving challenges of deepfakes, balancing individual rights, public interest, and the need for technological accountability.

An Extra Mile To Deal With The Deepfakes Dilemma

Addressing the multifaceted challenges posed by deepfakes in India requires a balanced techno-legal approach that integrates technical innovation with robust legal frameworks. While the legislature and judiciary have made notable strides in curbing deepfake-related issues, additional measures must be explored to achieve comprehensive solutions. Globally, organizations have developed advanced technologies to identify and counter the spread of deepfakes. For instance, HONOR has introduced an in-device system capable of analysing videos to determine their authenticity[63], while Google's SynthID[64] facilitates watermarking and detecting AI-generated content, enabling compliance with watermarking obligations in certain jurisdictions. Similarly, Intel's FakeCatcher[65] provides tools to assess whether media has been manipulated using deepfake technology. YouTube's AI disclosure policy, which provides automatic labelling of content created using its Dream Track or Dream Screen tools, serves as a commendable example of responsible labelling and transparency, potentially mitigating the challenges posed by deepfake technology[66]. Technical specifications and guidance issued by organizations like C2PA[67] play a crucial role in enhancing data provenance, enabling the detection, restriction, and tracking of deepfake sources, thereby empowering stakeholders to swiftly address and mitigate the issues they create.

MeitY has initiated projects to identify technical solutions to address deepfakes[68], and several patent applications concerning deepfake countermeasures are currently under review[69]. Authorities should actively evaluate these applications to foster homegrown technical solutions tailored to the Indian context. Furthermore, introducing common industry standards for devices used in creating deepfakes can enhance safety and transparency. This could include adoption of International Organization for Standardization[70] or Bureau of Indian Standards'[71] standards related to deepfake technologies, ensuring uniformity and accountability across platforms.

Proactive measures are equally essential. Regulators and authorities must vigilantly monitor the dissemination of deepfakes and provide timely clarifications to keep citizens informed about potential risks[72]. Public awareness is critical, and citizens need guidance on verifying information as well as recognizing deepfake content. Educational initiatives and advisories from bodies like CERT-In[73] can play a pivotal role in enhancing digital literacy and equipping individuals with the tools to discern authentic content from manipulated media.

Additional regulations or reinterpretation of current laws to comprehend the advancement of technology may be required to ensure comprehensive protection against privacy violations, defamation, and unauthorized content creation facilitated by malicious synthetic media.[74] Further, legal instruments may be strengthened through exploring the necessity of explicitly defining deepfakes, mandating traceability measures like watermarking, and enhancing platform accountability for detecting and removing harmful content. Complementing these with technological tools, proactive enforcement, and international collaboration can ensure effective regulation and

mitigation of misuse.

Ultimately, addressing deepfake challenges necessitates a collaborative approach involving technological advancements, regulatory oversight, and citizen empowerment. By harmonizing these elements, India can develop an effective framework to mitigate the risks posed by deepfake technologies while promoting safe and ethical digital practices.

Conclusion

Tackling the complex challenges posed by deepfakes necessitates a paradigm shift toward a techno-legal approach. While legislative, judicial, and regulatory measures have made significant progress, a purely legal framework is insufficient to address the dynamic and evolving nature of deepfake technologies. A holistic approach that integrates cutting-edge technical solutions with legal safeguards is imperative. Global innovations such as deepfake detection tools, watermarking technologies, and proactive regulatory measures provide valuable insights, but India must also foster indigenous solutions by encouraging innovation and setting industry standards. Moreover, public awareness and digital literacy are equally crucial to empower individuals to recognize and report deepfakes. By adopting a cohesive strategy that balances technology, law, and public engagement, India can build a resilient framework to counter the misuse of deepfakes and other new-age technologies, ensuring a safer digital environment and protecting societal values in the face of rapid technological advancements.