DEEPFAKES IN INDIA: LEGAL CHALLENGES AND REGULATORY IMPERATIVES

Ankita Sultania*

ABSTRACT

Deepfake technology, driven by artificial intelligence (AI) and algorithms like generative adversarial networks (GANs), creates synthetic media such as images, videos, and audio that mimic real individuals with alterations. In India, the misuse of deepfakes gained prominence during political campaigns in 2020, sparking controversy and underscoring the technology's potential for deception and misinformation. This technology raises significant concerns about privacy violations, misinformation, and cybercrime, necessitating robust legal frameworks. While India's Information Technology Act (IT Act) offers some recourse, specific legislation targeting deepfakes is lacking. This gap complicates legal prosecution against offenders using deepfakes for malicious purposes. The risks associated with deepfakes are profound, including social discord, polarisation, and threats to individual privacy and further undermine trust in media and institutions by casting doubt on genuine information.\ India's response includes guidelines urging media labelling of manipulated content, yet formal legislative measures are needed to effectively combat deepfakes. Mitigating these risks demands technological advancements, stringent legal measures, media literacy initiatives, and international cooperation. Addressing these challenges will safeguard individual rights, uphold digital integrity, and restore trust amidst the evolving landscape of AI-driven synthetic media.

Keywords: Deepfake, AI, Technology, Privacy.

INTRODUCTION

Deepfake technology, an application of artificial intelligence (AI), employs algorithms such as generative adversarial networks (GANs) to create synthetic media like images, videos, and audio. Its primary aim is to produce highly realistic fake content that mimics real individuals while introducing alterations. Deepfakes rely on two key processes: deep learning, a subset of AI that utilises artificial neural networks inspired by the human brain to analyse vast amounts of data, and GANs, a specific architecture where a generator creates synthetic samples and a

*BBA LLB, FOURTH YEAR, NMIMS KIRIT P. MEHTA SCHOOL OF LAW, MUMBAI.

discriminator evaluates their authenticity against real data from the training set. This iterative process continues until the generator can produce convincing synthetic content.

In 2020, the first notable use of AI-generated deepfakes in political campaigns emerged when manipulated videos of Bharatiya Janata Party (BJP) leader Manoj Tiwari and Madhya Pradesh Congress chief Kamal Nath circulated widely, causing significant controversy and highlighting the potential misuse of deepfake technology[1].

The proliferation of deepfakes raises concerns about the need for robust legal frameworks to address issues such as privacy, misinformation, and cybercrime. While India has some legal provisions under the Information Technology Act (IT Act) that can be applied, there is a clear necessity for more explicit regulations tailored to the unique challenges posed by deepfakes. One critical issue is the lack of precise legal definitions and actions regarding offences involving deepfakes, complicating the prosecution of individuals or entities engaging in malicious or deceptive activities using this technology.

GENERATING AND OPERATING DEEP FAKES

Deepfakes involve using AI algorithms and techniques like machine learning and photo editing to fabricate misleading imagery for disinformation purposes. These technologies, particularly GANs and machine learning, enable the creation of manipulated videos by mapping facial features from one source onto another person's face. Additionally, voice cloning technology is employed to replicate someone's voice accurately. While India acknowledges the need for stronger legislation to combat deepfake-related crimes and fraud, existing laws under various sections of the IT Act can provide both civil and criminal remedies. The government has issued warnings to online platforms to uphold their legal obligations to identify and promptly remove deceptive content[2].

Tools for generating deepfakes range from freely available software to sophisticated paid applications. These tools empower users, even those with limited technical expertise, to manipulate videos by replacing faces or altering voices. Online platforms offer cloud-based services for creating deepfakes, accessible via web browsers. GitHub hosts several open-source projects related to deepfake technology, providing resources for developers and researchers.

[1] Bhaumik A, “Regulating Deepfakes and Generative AI in India | Explained” (*The Hindu*, December 4, 2023)
[2] Sharma, Mridul, and Mandeep Kaur. "A review of Deepfake technology: an emerging AI threat." *Soft Computing for Security Applications: Proceedings of ICSCS 2021* (2022)

RISKS ASSOCIATED WITH ADVANCING DEEPFAKE TECHNOLOGY

Deepfake technology once considered a novelty, has evolved into a potent tool with serious implications for society, governance, and individual well-being. The misuse of deepfakes spans a spectrum of threats that range from undermining public trust to exacerbating social divisions and facilitating criminal activities[3].

Social Discord and Polarization

Deepfakes have transcended being mere novelties to become tools for spreading misinformation and disinformation on a massive scale. Malicious actors, including state-sponsored entities, extremist groups, and financial interests, can exploit deepfakes to disseminate false narratives. These false narratives can sow discord, amplify existing social tensions, and influence public opinion and political outcomes.

Pornographic deepfakes, specifically targeting women, pose significant psychological and financial risks. They not only violate personal privacy but also perpetuate harmful stereotypes and can lead to emotional distress, reputational damage, and even financial extortion.

Undermining Trust in Media and Institutions

The proliferation of deepfakes undermines trust in traditional media and authoritative institutions. By blurring the line between reality and fiction, deepfakes contribute to a culture of uncertainty where distinguishing between authentic and manipulated content becomes increasingly challenging.

Nation-states and non-state actors can weaponise deepfakes to destabilise public safety, sow confusion, and undermine trust in institutions and authorities. For instance, deepfakes can be used to depict political figures engaging in illegal or provocative actions, thus fueling public distrust and anti-government sentiments.

The "Liar's Dividend" Phenomenon

The "liar's dividend" refers to the phenomenon where genuine media and facts are dismissed as deepfakes or fake news. This phenomenon exploits the public's awareness of deepfake

[3] Gupta, Kashish. "The Future of Deepfakes: Need for Regulation." *Nat'l LU Delhi Stud. LJ* 5 (2023)

technology to cast doubt on verifiable information, eroding the credibility of reliable sources and fostering a climate of scepticism and misinformation.

Beyond visual content, advancements in AI allow for the cloning of voices and the creation of synthetic audio, enabling perpetrators to impersonate individuals and perpetrate financial scams. This poses a significant threat to consumer trust and financial security, affecting a wide swath of the population, including in India, where digital financial transactions are increasingly common.

While deepfake technology offers innovative possibilities, its misuse poses multifaceted risks to society. Addressing these risks requires a combination of technological advancements, robust legal frameworks, media literacy initiatives, and international cooperation to mitigate the harmful impacts of deepfakes on individuals, communities, and democratic institutions. As technology continues to evolve, proactive measures are essential to safeguarding public trust, protecting personal privacy, and upholding the integrity of information in the digital age.

INDIA'S RESPONSE TO TACKLING DEEPFAKES

Deepfake technology has found significant application in Indian politics, entertainment, and even cases of personal revenge. Instances such as the manipulation of political speeches or the creation of explicit content featuring public figures underscore the potential harm deepfakes can inflict on individuals and society as a whole[4].

Despite these challenges, India's current regulatory framework lacks specific provisions to effectively address deepfakes[5]. On January 9, 2023, the Ministry of Electronics and Information Technology (MeitY) issued guidelines to media organisations, urging caution in disseminating manipulated content and emphasising the importance of labelling such content clearly. MeitY also advised victims to file FIRs (First Information Reports) at local police stations and avail remedies under the IT rules.

Following public outcry over deepfake incidents involving public figures like Rashmika Mandanna, regulatory bodies and online platforms have been reminded of their responsibilities

[4] ETtech, "75% Indians Have Viewed Some Deepfake Content in Last 12 Months, Says McAfee Survey" *The Economic Times* (April 25, 2024)
[5] Jha, Piyush, and Simran Jain. "Detecting and Regulating Deepfakes in India: A Legal and Technological Conundrum." *Available at SSRN 4411227* (2021).

under Section 66D of the IT Act and Rule 3(1)(b) of the IT Rules. However, these warnings lack enforceability and do not constitute formal legislation.

Indian Prime Minister Narendra Modi has highlighted deepfakes as one of the greatest threats facing the nation, urging vigilance and responsible use of new technologies. He emphasised the need for public awareness about AI-generated content and its potential misuse.

According to the "State of Deepfakes" report by Home Security Legends, online deepfake videos have surged by 550%, reaching nearly 95,820 instances in 2023, with India identified as the sixth most vulnerable country to this emerging threat[6].

DEEPFAKES AND INDIAN LEGAL FRAMEWORK

In India, while there are no specific laws targeting deepfakes directly, existing legal provisions can be leveraged to address the various offences associated with their creation, distribution, and misuse.

Information Technology Act, 2000 (IT Act)

Sections 66D and 66E: Impersonation and Privacy Violations

Section 66D of the IT Act criminalises impersonation through computer resources, applicable to cases where deepfakes are used to misrepresent individuals. Section 66E addresses privacy violations by prohibiting the circulation of personal data or images without consent, encompassing scenarios where deepfake videos infringe on an individual's privacy.

Section 79: Intermediary Liability

Under Section 79, online platforms hosting deepfake content are considered intermediaries and are required to promptly remove objectionable content upon notification by an affected party or a court order. Recent judicial interpretations have reinforced this liability, emphasising the responsibility of platforms to mitigate harms caused by malicious deepfakes, such as copyright infringement.

[6] Sarkar, Diya, and Sudipta De Sarkar. "Combatting Deep-fakes in India–An Analysis of the Evolving Legal Paradigm and Its Challenges." (2024).

Sections 67A and 67B: Sexually Explicit Material

These sections impose stringent penalties for creating, publishing or distributing sexually explicit content and child sexual abuse material online. In response, platforms like Pornhub have implemented bans on deepfake pornographic content to comply with legal provisions aimed at preventing the dissemination of objectionable material.

CONSTITUTIONAL SAFEGUARDS

Article 21 - Right to Privacy

Protection of Personal Identity**:** Article 21 of the Indian Constitution guarantees the right to privacy, protecting individuals against unauthorised use of their identity or image in deepfake videos. Courts have expanded this right to include safeguards against the misuse of personal data and images, emphasising privacy as essential to human dignity and personal liberty.

Indian Penal Code (IPC)

Defamation (Section 499)

Individuals depicted falsely in deepfake videos can seek recourse under defamation laws, which protect against false statements that harm one's reputation.

Forgery (Sections 463 and 468)

Deepfakes created with the intent to deceive or defraud fall under forgery laws, which penalise the fraudulent creation or distribution of falsified electronic records or documents.

Sedition (Section 124A), Criminal Intimidation (Section 504), Voyeurism (Section 354C)

These sections provide additional legal avenues to prosecute individuals involved in creating or disseminating deepfakes that incite public disorder, threaten individuals, or invade personal privacy.

In summary, while specific legislation directly addressing deepfakes is lacking in India, the existing legal framework comprising the IT Act, constitutional protections, and relevant sections of the IPC provides a comprehensive basis for addressing offences related to deepfake

technology. Judicial interpretations continue to play a crucial role in defining the application of these laws in response to emerging challenges posed by advancements in digital manipulation and AI technologies.

RECENT JUDICIAL INTERVENTIONS

Recent legal cases highlight judicial concerns regarding the misuse of deepfake technology and its implications for privacy and public order. In cases like Chaitanya Rohilla v. Union of India, the Delhi High Court sought the government's response on regulating AI and deepfake technologies to prevent potential harms such as privacy violations and economic damage.

Similarly, celebrities like Anil Kapoor and Amitabh Bachchan have secured court injunctions against unauthorised use of their likenesses in deepfake content, emphasising the need for legal protections against commercial exploitation and reputational harm.

COUNTRIES' RESPONSES TO COMBAT DEEPFAKE THREATS

Several countries have implemented measures to address the growing threat posed by deep fakes:

The United Kingdom is set to introduce regulations requiring clear labelling of AI-generated photos and videos to enhance transparency and empower users to verify media authenticity.

The European Union's Digital Services Act (DSA) mandates social media platforms to adhere to labelling obligations, improve transparency, and assist users in verifying media authenticity.

South Korea has criminalised the distribution of deepfakes that endanger public interests, imposing penalties of up to five years in prison or fines.

China's cyberspace authority has mandated clear labelling of deepfakes to prevent public confusion and enhance media transparency.

The United States has established a task force under the Department of Homeland Security to combat digital content fraud, including deepfakes.

SUGGESTIONS AND CONCLUSION

Addressing the challenges posed by deepfakes requires a multifaceted approach involving legal, regulatory, and technological measures. India needs comprehensive legislation specifically tailored to address the unique threats posed by deepfake technology. The current

regulatory framework, while encompassing aspects of AI and digital offences, lacks specific provisions to effectively combat deepfakes.

Proposed measures include

Enhancing legal provisions under the IT Act to explicitly cover deepfake-related offences, including stringent penalties for perpetrators.

Implementing stricter guidelines for online platforms to proactively detect and remove deepfake content, with penalties for non-compliance.

Promoting public awareness campaigns to educate users about the risks associated with deepfakes and encourage responsible digital consumption.

Establishing a dedicated regulatory authority to monitor and enforce compliance with deepfake-related regulations.

Encouraging international cooperation to develop standardised approaches for combating deepfake threats across borders.

In conclusion, while deepfake technology presents numerous challenges, it also offers creative and beneficial application opportunities. However, effective regulation and oversight are essential to mitigate the potential harms associated with its misuse. By adopting a proactive and collaborative approach, India can safeguard individual privacy, uphold digital integrity, and foster trust in online media.