LEGAL CHALLENGES IN REGULATING DEEPFAKE TECHNOLOGY UNDER INDIAN CYBER LAW

Mehak Bisht[*]

ABSTRACT

The advent of deepfake technology, a fusion of "deep learning" and "fake," represents a significant milestone in artificial intelligence, enabling the creation of hyper-realistic synthetic media. While the technology offers transformative opportunities in sectors like healthcare, entertainment, and tourism, its misuse has profound implications for individual privacy, societal trust, and democracy. Deepfake-related challenges include the proliferation of misinformation, political manipulation, social destabilization, and legal dilemmas, such as defamation and intellectual property rights infringement. In India, existing legal frameworks, including the Indian Penal Code (IPC) and the Information Technology (IT) Act, partially address issues like identity theft and defamation but fail to encompass the unique complexities of deepfake technology. Key challenges include safeguarding privacy under Article 21 of the Indian Constitution, combating cyber fraud, and ensuring content moderation on digital platforms. In the Indian context, the threats posed by deepfakes are particularly acute due to the nation's vast internet user base, limited media literacy, and diverse socio-political landscape. Challenges include their use for misinformation campaigns, political manipulation, erosion of democratic integrity, and defamation. Notable incidents, such as deepfakes of prominent political figures during elections and morphed videos of celebrities, illustrate the potential for societal harm, including communal tensions and destabilization. Moreover, deepfakes strain journalistic credibility, as they complicate the differentiation between genuine and manipulated media. The paper underscores the need for a robust, nuanced legal framework in India, drawing inspiration from international models like the EU's Artificial Intelligence Act and U.S. legislation. By bridging legislative gaps and leveraging global best practices, India can balance technological innovation with the protection of individual rights and societal integrity.

[*] BA LLB, FIFTH YEAR, ICFAI LAW SCHOOL, ICFAI UNIVERSITY DEHRADUN.

Keywords: Deepfake, cyber law, Artificial Intelligence (AI), Privacy, Misinformation.

INTRODUCTION

The term “deepfake” is a portmanteau of *“deep learning”* and *“fake”*, referring to the AI-driven methods used to create synthetic media. The technology has advanced rapidly, making it increasingly difficult to distinguish between authentic and manipulated content, which poses significant challenges to legal and regulatory frameworks.[1] According to the definition provided by the Merriam-Webster Dictionary, “an image or recording that has been convincingly altered and manipulated to misrepresent someone as doing or saying something that was not actually done or said is called deepfake”.[2] It further states that a deepfake is an image, or a video or audio recording, that has been edited using an algorithm to replace the person in the original with someone else (especially a public figure) in a way that makes it look authentic.[3] Deepfake technology is based on an advanced algorithm that produces hyper-realistic videos using a person’s face, voice or likeness by utilizing machine learning techniques.[4] It can be used for both malignant and munificent purposes, depending on how it is wielded.

Face reenactment, face generation, face swapping, and speech synthesis are some of the various ways in which deepfake technology is used. In the world of the internet and social media, this technology is used by several celebrity fanbases, influencers and social media channels and accounts to develop creative content, whether it is posting some cover song in their favorite music artists’ voice or showing their favorite artist embodied in the culture of their community resulting in uniting and bridging the gap between various cultures around the world or posting creative content being it informative, humorous or just depicting the skills of the maker in creating a particular art. Moreover, the utilization and progression of deepfake technology as a medium for enhancing various industries, including healthcare and entertainment, holds vast potential.[5] Notably, in the healthcare sector, deepfake has demonstrated significant diagnostic advantages, as evidenced by studies conducted by several scholars, while in the entertainment

[1] S. Gupta and A. Jain, Deepfake Technology and its Legal Ramifications in India (2021), 15, Journal of Cyber Law, 18-102.
[2] “Deepfake” (Merriam-Webster Dictionary) https://www.merriam-webster.com/dictionary/deepfake
[3] Supra Note 1.
[4] Vig, Shinu. "Regulating Deepfakes: An Indian perspective." (2024) 17, no. 3, Journal of Strategic Security 70.
[5] Geraint Rees, “Here's How Deepfake Technology Can Actually Be a Good Thing” (World Economic Forum Agenda, 2019).

industry, it can be used for dubbing and language localization.[6] Additionally, in the tourism industry, the photorealistic imageries created by the use of technology are particularly valuable for marketing and advertising purposes.[7]

However, just like any other technological development throughout history, this double-edged technology can also be used for sinful purposes and to fulfil ulterior motives. Numerous cases of morphing, technological harassment and fraud and Intellectual Property Rights infringement emerge every day as a result of the misuse of deepfake technology.

In recent years, the emergence of deepfake technology has presented new challenges to legal systems around the world, and India is no exception. Deepfakes, which use artificial intelligence to create hyper-realistic fake content, have profound implications for individual privacy, reputation, and the broader societal trust in digital media. As technology becomes more sophisticated, the legal frameworks in India, which are often based on outdated concepts and traditional methods of media regulation, find it increasingly difficult to keep pace.

RAMIFICATIONS OF DEEPFAKE AND CHALLENGES POSED BY IT

The rapid development of deepfake technology has led to escalating concerns, primarily focused on the creation of an environment lacking in trust, as well as raising the possibility of exploitation of vulnerable sections of society by corrupt scammers and hackers.[8]

Deepfake technology presents significant challenges, blurring the lines between reality and illusion with critical implications for human rights and regulatory frameworks, and at the same time, introducing multifaceted threats, particularly it's capacity to distort truth, manipulate public opinion and infringe upon the dignity and privacy of individuals.[9] The following are some challenges posed by deepfakes:

[6] Mika Westerlund, "The Emergence of Deepfake Technology: A Review," (2019) 9, no. 11, Technology Innovation Management Review .

[7] Patricia Picazo and Sergio Moreno-Gil, "Analysis of the Projected Image of Tourism Destinations on Photographs: A Literature Review to Prepare for the Future," (2019) 25, no. 1, Journal of Vacation Marketing, 3-24.

[8] Supra Note 4, 72.

[9] Swati Malik, "Blurring Boundaries Between Truth and Illusion: Analysis of Human Rights and Regulatory Concerns Arising from Abuse of Deepfake Technology" (2024) AIP Publishing.

1. Misinformation and Fake News

Due to its hyper-realist and persuasive nature, deepfake technology can easily be used to manipulate collective recollection of any specific event among masses, hence spreading misinformation rapidly, often providing the victims neither the opportunity nor the time to deny or control the escalation of such information or clear their name. The easy accessibility of audio-visual content through social media, coupled with the accessibility to modern tools such as Tensorflow or Keras, AI software, and cost-effective computing resources, along with the swift advancement of deep-learning (DL) methods, specifically Generative Adversarial Networks (GAN), has facilitated the creation of deepfakes with the intention of disseminating disinformation, monetary frauds, hoaxes, and disrupting government operations.[10] As a result, the identification and detection of deepfakes play a critical role in enhancing the credibility of social media platforms and other media-sharing websites.[11]

2. Political Manipulation

There have been numerous occurrences worldwide where deepfakes have been employed to advance political interests by exploiting the likeness and image of renowned political figures.[12] These individuals may be portrayed as using a racial epithet, accepting a bribe, and admitting to complicity in a crime, among other things.[13] India, being a country having around 323 million internet users, coupled with the lack of media literacy among its masses, serves as a potential hub for being the disseminator of misinformation. For instance, the misuse of deepfake in order to spread misinformation was noticed to be especially prevalent during elections this year. Major cases of misuse hit the headlines in April including deepfakes of Bollywood actors criticizing Prime Minister Narendra Modi and fake clips involving two of Modi's top

[10] Momina Masood, Marriam Nawaz, Ali Javed, Tahira Nazir, Awais Mehmood, and Rabbia Mahum, "Classification of Deepfake Videos Using Pre-trained Convolutional Neural Networks," (IEEE, 2021), 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2) 1-6.

[11] Paarth Neekhara, Brian Dolhansky, Joanna Bitton, and Cristian Canton Ferrer, "Adversarial Threats to Deepfake Detection: A Practical Perspective" (2021) Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 923-932.

[12] Supra note 4, 7.

[13] Abdul-Rahman Kabbara. "Bots & Deepfakes" (August 2021) NSI Intern Integration Project, https://nsiteam.com/social/wpcontent/uploads/2021/08/IIJO_eIntern-IP_Bots-andDeepfakes_Kabbara_FINAL.pdf.

aides that led to the arrest of nine people.[14] However, this wasn't the first time that the country had witnessed such an occurrence. Even prior to this, India witnessed its first-ever use of AI-generated deepfake technology in political campaigning in 2020 when several deepfake videos of politician Manoj Tiwari were circulated on WhatsApp groups, depicting Tiwari making accusations towards his political rival Arvind Kejriwal in both English and Haryanvi languages, preceding the elections in Delhi, the Indian capital state.[15]

3. Democracy Concerns

In a diverse country like India, with people having different religions and societal backgrounds, it is natural for the population to have different political, religious and general opinions. People are sensitive in respect of several political and religious topics enhancing the potential of deepfakes to exacerbate existing communal tensions, incite violence, and erode trust in democratic institutions, leading to social unrest and instability.

Such deepfakes have a significant impact on journalism as well. Media houses might also unintentionally spread misinformation due to their inability to differentiate between real and fake leads, resulting in the loss of faith of the masses in the press which is often classified as the fourth pillar of democracy. Moreover, politicians may use real videos as a defence against accusations by claiming they are deepfakes, a tactic known as the 'liar's dividend'.[16] Additionally, countries with strained diplomatic relations or territorial disputes with India, such as China or Pakistan, or Non-State actors like terrorist organisations, may also seek to exploit deepfake technology to spread propaganda, incite violence, or undermine the stability of the country and influence public opinion through the dissemination of such deepfake content.[17] Hence, it has now become the need of the hour to build safeguards against the malign use of

[14] AI and Deepfakes: Unveiling the Dark Side of Election Campaigns in India, The Economic Times, https://economictimes.indiatimes.com/news/india/ai-and-deepfakes-unveiling-the-dark-side-of-election-campaigns-in-india/articleshow/110169142.cms?from=mdr (last visited 25th November, 2024).

[15] Supra note 4, 7.

[16] Robert Chesney and Danielle Citron, "Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics" (2019) 98 Foreign Affairs, 147.

[17] Supra note 4, 8.

deepfake in order to protect the integrity of the democracy of the country and to maintain its stability.

4. Defamation and Legal Framework under IPC:

Defamation is one of the most immediate concerns in the face of deepfakes. Under Indian law, defamation is a criminal offense under Section 499 of the Indian Penal Code (IPC). However, deepfakes create a new dimension to this offense. A deepfake video or image can distort an individual's words or actions, potentially causing significant harm to their reputation. Given that deepfake content can be rapidly shared across social media platforms, tracing the original creator or the individual responsible for distributing harmful content becomes exceedingly complex. Current laws are not equipped to address the speed and scale at which such content spreads, and new legal tools may be necessary to address the nuances of AI-generated defamation.

LEGAL FRAMEWORK IN INDIA

Right to Privacy

The right to privacy, upheld as a fundamental right under Article 21 of the Indian Constitution in *K.S. Puttaswamy v. Union of India* (2017)[18], emphasizes an individual's control over the dissemination of personal information. This includes protection against the unauthorized creation and distribution of deepfakes, which threaten personal autonomy and dignity. The Digital Personal Data Protection Bill, 2022 addresses data privacy concerns, including unauthorized data usage and dissemination, by proposing penalties for breaches and establishing a Data Protection Board for grievances.[19] Its potential to encompass deepfake misuse, however, remains untested as the Bill awaits passage.

Simultaneously, tensions exist between the Indian government and social media companies over privacy and regulatory compliance. The *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021*, mandate intermediaries to disclose user identities upon government requests and ensure accountability for harmful user-generated

[18] Justice K.S. Puttaswamy (Retd.) & Anr. vs. Union of India & Ors., (2017) 10 SCC 1, AIR 2017 SC 4161.
[19] Digital Personal Data Protection Bill 2022 (India) - Proposed legislation regulating personal data in India.

content.[20] Non-compliance risks losing intermediary status and legal protections.[21] Social media platforms argue these obligations overreach, jeopardizing user privacy and freedom of expression, while the government emphasizes security and public safety. Addressing these conflicts demands a balance between privacy, regulatory oversight, and freedom of expression through collaborative stakeholder engagement.

Information Technology Act, 2000

Section 66E of the Information Technology (IT) Act, 2000, applies to deepfake offenses involving the unauthorized capture, dissemination, or transmission of an individual's visual representations via mass media, thereby infringing their privacy rights. Violations under this section may result in imprisonment of up to three years or a fine of ₹2 lakh.[22] Similarly, Section 66D of the Act addresses cyber fraud, including impersonation or deception through digital communication, with penalties of up to three years of imprisonment and/or a fine of ₹1 lakh.[23] These provisions serve as the legal basis for prosecuting deepfake-related cyber crimes in India.

Deepfakes can also be used for spreading disinformation, undermining government authority, or inciting public unrest. Such activities fall under Section 66F of the IT Act, 2000, which pertains to cyber terrorism, as well as the amended Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2022.[24] Hate speech and defamation facilitated by deepfakes pose significant risks to individual reputation and societal well-being, contributing to a toxic online environment.

Intermediaries hosting deepfake content may also be held accountable under Section 79 of the IT Act. This provision mandates that platforms remove unlawful content upon gaining knowledge of its presence or receiving a judicial order. However, in *Myspace Inc. v. Super Cassettes Industries Ltd.* (2017)[25], the Court clarified that intermediaries must remove infringing content upon receiving notice, even in the absence of a court directive. Additionally, under the IT Rules of 2021, significant social media intermediaries are required to appoint

[20] *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021* (India)- Guidelines for intermediaries and digital platforms introduced by the Indian government.

[21] Ibid.

[22] Information Technology (IT) Act, 2000, Section 66E (India): Governs privacy violations related to visual representations shared without consent.

[23] IT Act 2000, Section 66D (India): Addresses cyber fraud and impersonation.

[24] IT Act 2000, Section 66F (India): Defines cyber terrorism and its ramifications for public safety.

[25] *Myspace Inc. v. Super Cassettes Industries Ltd.*, (2017) 236 DLT 478 (DB)- Established intermediary obligations for removing unlawful content upon private notice.

grievance officers and implement mechanisms to address complaints regarding harmful content.[26]

In November 2023, the Indian government issued a directive mandating social media platforms to remove deepfake videos within 24 hours of a complaint. This followed incidents involving morphed videos of actors, emphasizing the urgency to regulate deepfake proliferation effectively.[27]

Copyright Laws

If an individual utilizes copyrighted material without permission to create deepfakes, copyright holders have the right to initiate legal proceedings against the infringing party. The Indian Copyright Act of 1957, specifically Section 51, establishes penalties for specific offenses related to copyright infringement, explicitly prohibiting the unauthorized utilization of another individual's property, particularly if the said property is subject to exclusive ownership.[28] Deepfakes, which encompass the unsanctioned manipulation or modification of extant photos and videos, may be deemed to encroach upon the copyright proprietor's entitlement to reproduction of the copyrighted material in the event that a significant segment is replicated.[29] By virtue of deepfakes being creations founded upon pre-existing works, these can also be regarded as derivative works.[30] As copyright law confers the exclusive right to produce derivative works solely upon the copyright owner, the act of creating and disseminating deepfakes without authorization would be deemed as copyright infringement.

POSSIBLE SOLUTIONS

Globally, countries like the United States and the European Union have already begun to implement regulations addressing deepfakes. The U.S. has enacted the Malicious Deep Fake Accountability Act of 2018, while the EU has included deepfake regulation in its Artificial Intelligence Act.[31] These international efforts offer useful insights for India’s legal system, particularly in terms of creating a legal framework that balances free speech, privacy, and the

[26] IT Rules 2021 (India): Specifies intermediary accountability mechanisms, including grievance redressal and content monitoring.
[27] Indian Government Directive, November 2023, under the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 (India).
[28] Copyright Act 1957, Section 51 (India)- Specifies infringement penalties for unauthorized use of copyrighted material.
[29] Supra note 4.
[30] Ibid.
[31] European Union General Data Protection Regulation (GDPR).

prevention of harm caused by synthetic media.[32] Taking inspiration from the same, we can introduce the following measures in order to prevent the misuse of deepfake technology in the country:

1. Separate Legislation: Pass a separate law dealing with manipulation relating to deepfake; to provide measures regarding the qualities of deep fake, the kinds and definitions of deep fake, and civil and penal justice systems regarding the defense against deepfake offense and methods of recognizing and managing them.

1. **Establishment of Specialized Research Institutions**: Create dedicated institutions focused on the study of deepfake technology, encompassing its generation, applications, and potential misuse. These institutions should also work on developing advanced countermeasures and detection tools.

2. **Stringent Legal Frameworks and Penalties**: Impose severe penalties on individuals involved in creating, disseminating, or unlawfully possessing deepfake content. Such measures should act as a deterrent to potential offenders.

3. **Victim-Centric Remedies**: Provide comprehensive remedies for victims of deepfake abuse, including options for immediate takedown of the content, rehabilitation measures, and compensatory damages.

4. **Streamlined Complaint Mechanisms**: Develop user-friendly and accessible processes for filing complaints and initiating legal action against unauthorized deepfake material, ensuring prompt resolution.

5. **Protection of Victim Confidentiality**: Mandate the safeguarding of victims' identities throughout investigations and judicial processes. This responsibility should rest with law enforcement and judicial authorities to encourage reporting without fear of stigma or reprisal.

6. **Creation of Regulatory and Enforcement Bodies**: Establish new regulatory agencies tasked with identifying, monitoring, and penalizing offenders involved in deepfake-

[32] Civilsdaily, “Regulating Deepfake Technology under Indian Cyber Law” (2023).

related offenses. These bodies should operate in coordination with existing law enforcement and judicial systems to enhance accountability and efficiency.

7. **Public Awareness and Education:** Participate in public deliberate communication which will entail; carrying out campaigns and education programs to explain to people the risks of falling victim to deep fakes; secondly raise the people's media literacy and competency level to between the original fakes and the more manufactured ones.

CONCLUSION

The regulation of deepfake technology under Indian cyber law poses multifaceted challenges that demand urgent and targeted intervention. While existing legal frameworks, such as the Indian Penal Code (IPC) and the Information Technology (IT) Act, 2000, address some aspects like defamation and identity theft, they fall short in tackling the unique and complex nature of AI-generated deepfakes. The rapid evolution of this technology has outpaced traditional legislative and enforcement mechanisms, creating significant gaps in areas such as privacy protection, intellectual property rights, and the mitigation of misinformation.

India's recognition of the right to privacy as a fundamental right through *K.S. Puttaswamy v. Union of India* and the proposed Personal Data Protection Bill, 2019, reflect progress toward safeguarding individual rights in the digital era. However, the absence of specific provisions addressing deepfakes highlights the need for a more nuanced legal framework. Lessons can be drawn from global approaches, such as the EU's Artificial Intelligence Act and the United States' legislative efforts targeting malicious deepfakes.

Moving forward, Indian lawmakers must prioritize the formulation of robust laws that comprehensively address the creation, distribution, and malicious use of deepfakes. This includes stricter content regulation policies, clearer accountability mechanisms for creators and platforms, and improved technological tools for law enforcement to identify and mitigate the spread of synthetic media. By fostering international cooperation and leveraging global best practices, India can build a legal infrastructure that balances technological innovation with the protection of individual rights and societal integrity.