NAVIGATING THE LEGAL LANDSCAPE: ADDRESSING DEEPFAKE CONCERNS IN INDIA THROUGH ENHANCED LEGISLATIVE FRAMEWORKS AND COLLABORATIVE STRATEGIES

Rishita Yadav*

ABSTRACT

In the era of informational capitalism, social media applications have been converted into platforms where individuals, through forms of self-expression, generate content for these websites and thus contribute significantly to their overall revenue. With the coming of Artificial Intelligence (AI), technology has advanced a step further and a concept called "Deepfake" has come into the picture. While deepfakes can be harnessed positively, such as providing a new voice for those who have lost their speech as well as enhancing entertainment quality, the negative consequences cannot be ignored. Individual privacy has been exploited to a great extent by the fabrication of falsified videos, fake images, and manipulated audio with the help of AI which in turn has serious consequences ranging from reputational damage to financial losses and psychological distress. In November 2023, a video of actor Rashmika Mandanna went viral on social media and it severely affected her reputation with visceral reactions from both the celebrity and the public. Her experience depicts the vulnerability of individuals on social media platforms who fall prey to content-hungry people, who disseminate falsified and impersonated content to gain widespread attention and thus result in large monetary benefits for these people. In response to this, social networking sites have also received notices from the Indian IT ministry informing them that impersonating someone online is prohibited according to Section 66D[1] of the Information Technology Act of 2000. The existing laws in India concerning deepfakes are insufficient to comprehensively tackle the problem. The absence of explicit clauses related to artificial intelligence, machine learning, and deepfakes in the IT Act of 2000[2] creates challenges in efficiently governing the misuse of these technologies. The research critically evaluates the existing legal frameworks in India to tackle deepfake-related offences, emphasizing the need for adaptive legislation to address the rapidly evolving nature of AI technology.

*BBA LLB, SECOND YEAR, SYMBIOSIS LAW SCHOOL, NOIDA.

[1] Information Technology Act 2000, s 66D

[2] Information Technology Act 2000

Keywords: Artificial Intelligence, Deepfakes, Privacy, Informational Capitalism, Falsified Videos, Fake Images, Manipulated Audio, Adaptive Legislation.

WHAT IS DEEPFAKE?

Deepfake is a type of artificial intelligence technology that creates synthetic media such as photos, videos, and sounds by using machine learning techniques, most especially generative adversarial networks (GANs). Deepfake technology's main goal is to create very realistic synthetic content that looks like actual people, but with some aspects changed. Two fundamental methods that form the basis of deepfake technology are generative adversarial networks and deep learning.[3]

Artificial neural networks, a type of algorithmic framework inspired by the composition and operations of the human brain, are used in deep learning, a branch of machine learning. Large datasets may be processed and analysed using this method, which has been effectively employed in several domains including robotics, computer vision, natural language processing, and speech recognition.

One particular deep learning architecture is represented by generative adversarial networks or GANs. To create fresh synthetic data that closely mimics the original data, GANs train two neural networks—a discriminator and a generator—using a dataset.[4] The discriminator assesses the veracity of the created samples as well as the genuine ones from the training dataset during this process, while the generator creates fictitious examples. An adversarial training process is used, with the discriminator trying to correctly identify between the produced and genuine samples and the generator trying to produce examples that fool the discriminator. Until the generator can create stunningly realistic synthetic data, this iterative procedure is continued.

The generated deep fakes can be used in different ways for both video and image manipulations, such as:

(a) **face swap**, where one person's face is replaced with another's;

[3] Shubham Pandey and Gaurav Jadhav, 'Emerging Technologies and Law: Legal Status of Tackling Crimes Relating to Deepfakes in India' (*SCC Blog*, 17 March 2023) <https://www.scconline.com/blog/post/2023/03/17/emerging-technologies-and-law-legal-status-of-tackling-crimes-relating-to-deepfakes-in-india/> accessed 26 January 2024

[4] *Ibid*

(b) **attribute editing**, allowing changes to the person's characteristics, like altering hair colour or style;

(c) **face re-enactment**, transferring facial expressions from one person to another in the target video; and

(d) **fully synthetic material**, where real data is used for training, but the resulting picture is entirely fabricated.

DEEPFAKES- THE PRESENT SCENARIO

The first known application of deepfake technology is said to have occurred in 2017 when a Reddit user utilised a publicly accessible AI-driven programme to produce pornographic content by superimposing the faces of celebrities on the bodies of regular people.[5] Editing tools, such as Photoshop, have been around for decades. Deepfakes may now be created by inexperienced or semi-skilled people, making it possible to manipulate pictures and audio-visual recordings. In 2020, the Deeptrust Alliance—a confederation of business and civil society stakeholders—released a warning statement pointing out that disinformation-creation and -spreading technologies are now more readily available, affordable, quicker, and simpler than before.[6]

Resources to empower people against the exploitation of deepfakes and related technologies are becoming more widely available as their detection gets more difficult. For example, the Massachusetts Institute of Technology (MIT) created the Detect Fakes website, which focuses on minute features to help users recognise deepfakes.

There are serious worries about the use of deepfakes to enable technology-driven online gendered violence. In a 2019 research, the AI company Deeptrace discovered that an astounding 96% of deepfakes were determined to be sexual, with women appearing in 99% of them.[7] The Internet Freedom Foundation (IFF) founding director, Apar Gupta, is a lawyer who emphasises how deepfake technology is weaponized against women. He cites examples of

[5] Aaratrika Bhaumik, 'Regulating Deepfakes and Generative AI in India: Explained' (*The Hindu*, 4 December 2023) <https://www.thehindu.com/news/national/regulating-deepfakes-generative-ai-in-india-explained/article67591640.ece> accessed 29 January 2024

[6] *Ibid*

[7] Tom Simonite, 'Most Deepfakes Are Porn, and They're Multiplying Fast' (*Wired*, 7 October 2019) <https://www.wired.com/story/most-deepfakes-porn-multiplying-fast/> accessed 29 January 2024

romantic partners using deepfakes to shame women for rejecting their advances, resulting in social repercussions as well as psychological trauma.

LEGAL FRAMEWORK REGULATING THE ISSUE OF DEEP FAKES IN INDIA

Privacy Violation under IT Act: Section 66E[8] of the Information Technology Act, 2000 (IT Act) has provisions about deepfake crimes, which are defined as crimes when someone's photographs are captured, published, or transmitted in mass media, resulting in infringements on their privacy. The punishment is up to three years of imprisonment or a fine of ₹2 lakh.

Impersonation and Cheating under the IT Act: Section 66D[9] of the IT Act imposes penalties on individuals who employ communication devices or computer resources with malicious intent, resulting in impersonation or cheating facilitated by the use of deepfake technology. Punishment for this can be up to three years imprisonment and/or a fine of ₹1 lakh.

Obscene Content under IT Act: Sections 67, 67A, and 67B of the IT Act[10] give legal authority to prosecute individuals who spread deepfake content that consists of obscene or explicit sexual materials by sharing. Social media platforms that do not quickly remove 'artificially morphed images' run the risk of losing their ‘safe harbour’ protection.

IPC Provisions for Cybercrimes: Provisions such as Section 509 (offence related to insulting a woman's modesty), Section 499 (criminal defamation), and Section 153 (a) and (b) (inciting hatred on communal grounds) in the Indian Penal Code, 1860,[11] can be applied to address cybercrimes associated with deepfakes.

Copyright Act for Unauthorized Use: The Copyright Act of 1957[12] can be invoked if copyrighted material is used in the creation of deepfakes. Section 51[13] of the Copyright Act makes it illegal to use someone else's stuff without permission if they have exclusive rights to it. This gives a legal way to address copyright infringement issues related to deepfakes.

In addition to these legal actions, media organisations were advised by the **Ministry of Information and Broadcasting** on January 9, 2023 to exercise caution while airing anything

[8] Information Technology Act 2000, s 66E
[9] Information Technology Act 2000, s 66D
[10] Information Technology Act 2000, s 67, 67A and 67B
[11] The Indian Penal Code 1860, s 509, 409 and 153
[12] The Copyright Act 1957
[13] The Copyright Act 1957, s 51

that may be manipulated or tampered with. Additionally, the Ministry advised media outlets to mark any edited content as "manipulated" or "modified" to alert viewers to its altered state.[14]

The Ministry of Electronics and Information Technology, in its most recent Advisory dated November 07, 2023, has issued directives to prominent social media intermediaries that include:[15]

- Make sure to diligently and reasonably identify misinformation and deepfakes, especially information that breaches rules, regulations, or user agreements.
- Act promptly on such instances, adhering to the specified timeframes outlined in the IT Rules 2021.
- Users are advised against hosting such information or content, including Deep Fakes. If reported, promptly remove such content within 36 hours of the report.
- Take swift action by the timelines specified in the IT Rules 2021, and disable access to the content or information.

INDIAN CASES INCLUDING DEEP FAKES

Anil Kapoor's Lawsuit:[16]

- Bollywood actor Anil Kapoor took legal action against the use of AI-generated deepfakes featuring his likeness and voice.
- The deepfake content included GIFs, emojis, ringtones, and explicit material.
- The Delhi High Court granted protection to his attributes, issuing a court order to stop sixteen entities from using his name, image, or likeness through AI for financial gain.

Amitabh Bachchan's Case:[17]

- Actor Amitabh Bachchan also faced a situation where his personality rights, including voice, name, image, and likeness, were being used for commercial purposes without authorisation.

[14] Gandhi A, Rana V and Thakur R, 'Deepfakes and Breach of Personal Data – a Bigger Picture' (*Live Law*, 24 November 2023) <https://www.livelaw.in/law-firms/law-firm-articles-/deepfakes-personal-data-artificial-intelligence-machine-learning-ministry-of-electronics-and-information-technology-information-technology-act-242916?infinitescroll=1> accessed 29 January 2024

[15] *Ibid*

[16] *Anil Kapoor vs Simply Life India & Ors* (2023) 2023 LiveLaw (Del) 857

[17] *Amitabh Bachchan v. Rajat Negi and Ors.* (2022) SCC OnLine Del 4110

- In his case, the court granted an interim injunction, preventing the unauthorized commercial use of his attributes.

Rana Ayyub's Case:[18]

- Revenge Pornography Issue: Journalist and human rights activist Rana Ayyub faced a case highlighting the inadequacy of current laws in protecting individuals from revenge pornography. She was the victim of deepfake and her face was used in videos showing pornographic content. As a result, she received death and rape threats and her mental health situation was so disturbed that she could hardly eat or speak.
- Manipulated Explicit Video: An individual manipulated an explicit video to falsely portray Rana Ayyub, following her advocacy for a rape victim in Kathua.
- Online Harassment and Twitter Response: Rana Ayyub faced online harassment and hate speech. To combat this, she used Twitter, where manipulated tweets featuring her were being shared, adding to the challenges she faced.

The legal system in India fell short of addressing the issue of deepfake deeply since no express provision on it is available yet.

INTERNATIONAL BEST PRACTICES

United States: In the United States, an executive order on AI was signed by President Joe Biden according to which the Department of Commerce had the responsibility of labelling AI-created content on websites by "watermarking" so that it can easily be detected and appropriate measures can be taken at the earliest. The states of California and Texas have already implemented laws that criminalize the publication or distribution of videos or photos made due to deepfake technology that has the power to influence election outcomes.[19] Non-consensual distribution or publication of pornography invites criminal penalties in Virginia.[20]Moreover, the DEEP FAKES Accountability Bill, 2023[21] recently introduced in Congress puts a requirement on content creators or distributors to label AI-generated content explicitly and also

[18] *Rana Ayyub vs Directorate Of Enforcement* (2023) WP(C) 714/2022

[19] Titus Wu, 'California Looks to Boost Deepfake Protections before Elections' (*Bloomberg Law*, 15 December 2023) <https://news.bloomberglaw.com/artificial-intelligence/california-looks-to-boost-deepfake-protections-before-elections> accessed 29 January 2024

[20] Ruiz D, 'Deepfakes Laws and Proposals Flood Us: Malwarebytes Labs' (*Malwarebytes*, 22 January 2020) <https://www.malwarebytes.com/blog/news/2020/01/deepfakes-laws-and-proposals-flood-us> accessed 29 January 2024

[21] DEEPFAKES Accountability Act 2023

put notifications that alert the users regarding alterations in the videos, images or any other content. Failing to abide by these requirements would invite criminal penalties.

China: To control Deep synthesis technology and the spread of misinformation, the Cyberspace Administration of China published new guidelines to curb such practices. These guidelines ensure that any content created as a result of this technology can be traced back to its source. People providing such services using deep synthesis technology are supposed to abide by local laws, follow ethics and maintain the appropriate public opinion and political direction.

European Union: The EU has laid down the Code of Practice on Disinformation that ensures that big social media giants like Twitter, Google, Meta, etc. take appropriate measures to curb deepfakes or they would invite severe fines which could hinder their operational efficiency. Moreover, this code works hand in hand with the Digital Services Act[22] whose main aim is to keep a close check on these platforms so that any kind of misuse or breach can be prevented. Also, an act called the EU AI Act has been proposed which would ensure increased transparency, disclosure and accountability since deepfake providers would be subject to various strict requirements.

South Korea: South Korea has tackled the situation by putting up penalties and fines. According to its law, offenders who are engaged in the illegal distribution of deep fakes can face up to five years of imprisonment or fines that can be extended up to 50 million won.

SUGGESTIONS

The Information Technology Act of 2000[23], India's current cybercrime legislation, is unable to adequately address the growing threat posed by deepfakes. Since deepfakes, machine learning, and artificial intelligence are not particularly addressed in the IT Act, the Act must be amended to include laws that target deepfake usage and set forth repercussions for such abuse. This change should include enhancing the legal safeguards for people whose photos are used without their permission and stiffening the penalties for making and disseminating deepfakes with malevolent intent.

[22] Digital Services Act 2022
[23] Information Technology Act 2000

Recognizing the global nature of the deepfake challenge, India needs to engage in collaborative efforts with other nations. Multilateral initiatives can establish standardized regulations to govern the creation and usage of deepfakes, thereby ensuring a comprehensive and coordinated approach to privacy protection and cybersecurity. The government should take inspiration from the US, China and the EU's deepfake combating framework and embed it in its legislation. For instance, labelling or watermarking techniques followed in the US can be used so that the source can be traced at the earliest in cases of breaches or privacy violations. The government can also encourage its institutions like the IITs to come up with technology that can be used to detect deepfakes and invest in the same so that this growing problem can be addressed at the earliest.

In the meantime, the government could adopt specific measures to address the immediate threat. These include:

Censorship Strategy: A censorship strategy's implementation entails preventing publishers and middlemen from distributing misleading information to the general public. This strategy attempts to stop the distribution of deceptive material by limiting the channels through which deepfakes may be shared.

Legal Accountability: There should be an express provision in the current statutes like the IT Act that addresses the issue of deepfakes more effectively. Also, the punishments and fines for specific offences that can be committed using deepfakes should be provided in such statutes.

Regulation of Intermediaries: Using certain sections of the IT Act, the government can push online platforms to act fast in removing fake content. This means these platforms have to step up and work with authorities to stop the misuse of deepfakes.

CONCLUSION

In conclusion, the emergence of deepfake technology poses a serious threat to the legal system and highlights the necessity for comprehensive legislation to address its effects. The advent of remarkably lifelike altered media necessitates a revaluation of the existing legal structures intended to protect people from damage, as they contend with the rapidly developing field of digital deception. Deep fakes provide serious ethical and legal issues due to the blurry line separating fact from fiction. The potential for this technology to be abused for deceptive ends, such as disseminating false information or damaging reputations, highlights the need to take legislative action. Legislators must carefully strike a balance between protecting the right to free speech and reducing the dangers of misleading information.

The legal reaction to deep fakes should go beyond traditional borders, addressing civil, criminal, and technological issues. Civil liability rules must be strengthened to give meaningful remedies for victims of deep false manipulation. In a digital universe where reality is flexible, courts must punish the guilty and claim damages for the damaged party. Criminalising the development and spread of deep fakes is important for restricting potential offenders, and law enforcement authorities must have up-to-date tools and experience to deal with technologically advanced crimes. Collaboration among governments, technological corporations, and international organisations is critical for developing tactics that can adapt to the continuously changing nature of deep fake technology.

It is essential to integrate advanced detection algorithms and authentication mechanisms into the legal framework. Taking proactive steps is crucial to anticipate the challenges presented by the widespread and rapid dissemination of deep fakes, as coping strategies may prove inadequate. Additionally, there is a need for educational initiatives to raise awareness about the presence and potential dangers of deep fakes. Empowering individuals to critically assess media content is fundamental for building a resilient society capable of distinguishing between authentic and manipulated information. In summary, effectively addressing the convergence of deep fake technology and legal frameworks requires a comprehensive and adaptable approach. Collaboration among legislators, law enforcement, and technology experts is necessary to develop frameworks that efficiently reduce the risks associated with deep fakes. The legal response should be forward-thinking, covering civil remedies, criminal consequences, and technological advancements to uphold principles of truth, accountability, and justice.