HARMFUL USE OF DEEPFAKE TECHNOLOGY: A LEGAL AND DOCTRINAL ANALYSIS

Sanjana Sharma, Amity University

ABSTRACT

Deepfake technology, powered by artificial intelligence and machine learning, has emerged as one of the most disruptive technological advancements of the digital era. Although it offers legitimate applications in entertainment, education, and accessibility, its misuse has raised serious legal, ethical, and societal concerns. Deepfakes pose more than just misinformation, according to cyberlaw experts. It violates fundamental personality rights such as privacy, reputation, autonomy, and consent. Legal professionals and digital citizens face the difficulty of not just establishing authenticity, but also claiming ownership of their digital identity.

This paper examines the harmful use of deepfake technology with a specific focus on India while incorporating comparative global perspectives. It analyses instances of misuse such as political misinformation, non-consensual intimate imagery, financial fraud, and identity manipulation. Further, it explores judicial responses and emerging case law, highlighting the inadequacy of existing legal frameworks in addressing deepfake-related harms. Finally, it proposes regulatory reforms, including the need for a dedicated legal framework, platform accountability, and technological safeguards. The study concludes that while existing laws provide partial remedies, a comprehensive and forward-looking regulatory approach is essential to mitigate the risks posed by deepfake technology.

1. Introduction

The rapid advancement of artificial intelligence has drastically altered the creation and dissemination of digital content. Among these innovations, deepfake technology has emerged as both a powerful tool and a significant threat. The word "deepfake" is a portmanteau of "deep learning" and "fake" (Rouse, 2020). By leveraging techniques such as Generative Adversarial Networks (GANs), deepfakes enable the creation of hyper-realistic content that is often indistinguishable from authentic material. One of the most amazing features of its algorithmic design is its ability to create a video clip of a person speaking or acting in ways that they would never do or say in real life from just one picture of them[1].

Until late 2017, this machine learning technique was largely used for AI research. Only when a Reddit user known as "Deepfakes" began releasing digitally manipulated pornographic videos in which celebrities' faces were placed onto the bodies of women in pornographic films did this technology become widely known in the public domain (Schwartz, 2018). By the time Reddit later prohibited the posting and dissemination of deepfakes on its site, the creators of the videos had developed "FakeApp," an easy-to-use platform for creating fake media. Deepfake technology became well recognised and accessible to the general public thanks to FakeApp, resulting in a rapid increase in the number of people using this technology to make and disseminate deepfakes online, primarily through social media platforms. In September 2019, Deeptrace, an AI business, discovered around 15,000 deepfake videos online, 96% of which were pornographic. The harmful potential of deepfakes lies in their ability to distort reality, manipulate public perception, and infringe upon individual rights. Unlike traditional forms of misinformation, deepfakes carry a heightened degree of credibility due to their visual and auditory realism. This has profound implications for democratic processes, personal privacy, and national security.

In India, the proliferation of deepfakes has coincided with increased internet penetration and widespread use of social media platforms. With over 850 million internet users, the scale of potential harm is unprecedented[2]. Deepfakes have been used to spread political misinformation, create non-consensual sexual content, and perpetrate financial fraud. The increasing sophistication of these technologies poses significant challenges for law enforcement and

[1] Libby, K. (2019, August 13). This Bill Hader Deepfake Video Is Amazing. It's Also Terrifying for Our Future. Retrieved from https://www.popularmechanics.com/technology/security/a28691128/deepfake-technology/
[2] National e-Governance Division, *Deepfakes in India: Legal Landscape* (2025). (NeGD)

regulatory authorities. Globally, similar concerns have been raised. Instances of deepfake misuse in elections, financial scams, and cyber harassment highlight the transnational nature of the problem. Despite these challenges, legal systems across jurisdictions remain ill-equipped to address the unique harms posed by deepfakes.

This paper adopts a doctrinal approach to analyse the harmful use of deepfake technology, focusing on instances of misuse, legal responses, criminal liability, and the need for regulatory reforms.

2. The Dangers of Deepfakes

Due to the extensive use of deepfake technology in pornography and other areas, numerous scholars, especially in the fields of law and policy, have raised alarms about the possibility of deepfakes being used as powerful instruments for exploitation and sabotage, as well as their potential to negatively impact society by undermining democratic discussions on vital policy matters (Libby, 2019). In addition, recent occurrences, like Russia's meddling in the 2016 presidential election, along with the increasing threat of cyberwar escalation, have elevated the importance of evaluating the risks posed by deepfakes on the agendas of many organizations, including some of the largest social media platforms, such as Twitter and Facebook.

2.1 Understanding Deepfake Technology

It refers to the use of artificial intelligence (AI) techniques to create or manipulate audio-visual content in a manner that makes it appear authentic, despite being wholly or partially fabricated. The term “deepfake” is derived from “deep learning” and “fake,” reflecting the underlying use of advanced neural networks to generate synthetic media.[3] Over the past decade, rapid developments in machine learning have significantly enhanced the realism and accessibility of such technology, making it both a powerful tool and a potential instrument of harm.

At its core, deepfake technology relies on deep learning models, like Generative Adversarial Networks (GANs). GANs operate through a dual-network system comprising a “generator” and a “discriminator.”[4] The generator creates synthetic content, while the discriminator

[3] Chesney R and Citron D, ‘Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security’ (2019) 107 *California Law Review* 1753.
[4] Goodfellow I et al., ‘Generative Adversarial Nets’ (2014) *Advances in Neural Information Processing Systems* 2672.

evaluates its authenticity against real data. Through iterative training, the generator progressively improves its outputs until the fabricated content becomes nearly indistinguishable from genuine media. This process allows for the seamless manipulation of facial expressions, voice patterns, and even body movements.

Deepfakes can be broadly categorised into several types based on their mode of manipulation. First, face-swapping deepfakes involve replacing one person's face with another in a video, often without detectable distortion.[5] Second, voice cloning technologies replicate an individual's speech patterns, tone, and accent to produce convincing audio recordings. Third, full-body synthesis and text-to-video generation represent more advanced forms, enabling the creation of entirely artificial human personas or events that never occurred. These developments demonstrate that deepfakes are no longer limited to visual manipulation but extend to comprehensive identity simulation.

2.2 Deepfake Harms And The Liar's Dividend

While deepfake technology has legitimate applications, its misuse raises significant legal concerns. In the entertainment industry, for instance, it is used for visual effects, dubbing, and digital resurrection of actors. Similarly, in education and accessibility, deepfakes can assist in language translation and assistive communication. However, the same technology, when deployed maliciously, can distort reality, facilitate fraud, and infringe upon fundamental rights.

A key concern associated with deepfakes is their ability to undermine the epistemic trust traditionally associated with audio-visual evidence. Historically, photographs and videos have been regarded as reliable forms of proof in both legal and social contexts. Deepfake technology disrupts this assumption by introducing a new category of content that appears authentic but lacks any factual basis. This phenomenon has been described as the "liar's dividend," where the existence of deepfakes allows individuals to dismiss genuine evidence as fabricated, thereby complicating the pursuit of truth and accountability.[6] Epistemic uncertainty reduces trust in visual proof, lowers responsibility, and fuels cynicism. Deepfakes not only injure individuals, but also undermine confidence and justice in society.

[5] Tolosana R et al., 'DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection' (2020) 64 *Information Fusion* 131.
[6] Westerlund M, 'The Emergence of Deepfake Technology: A Review' (2019) 9 *Technology Innovation Management Review* 40.

From a legal perspective, deepfakes represent a qualitative shift in the nature of digital harm. Unlike conventional forms of misinformation, deepfakes are synthetic in origin, meaning they do not merely distort existing facts but create entirely new realities. This has implications for doctrines relating to privacy, defamation, and identity. For instance, the unauthorised use of an individual's likeness in a deepfake may constitute a violation of informational privacy, as recognised in K.S. Puttaswamy v. Union of India.[7] Similarly, the fabrication of false narratives through deepfakes can severely damage reputation, engaging principles articulated in R. Rajagopal v. State of Tamil Nadu.[8]

2.3 Democratisation of Deepfake Tools

Another critical dimension is the democratisation of deepfake tools. Earlier, the creation of high-quality manipulated media required significant technical expertise and resources. However, the proliferation of user-friendly applications and open-source software has made deepfake generation accessible to the general public.[9] This widespread accessibility increases the risk of misuse, as individuals with minimal technical knowledge can produce convincing fake content.

Furthermore, the rapid dissemination of deepfakes through social media platforms amplifies their impact. The virality of digital content ensures that deepfakes can reach large audiences within a short span of time, often before they can be detected or removed.[10] This creates a regulatory challenge, as the speed of technological misuse outpaces the capacity of legal and institutional responses.

Deepfake technology is not merely a technological phenomenon but a socio-legal challenge that intersects with issues of privacy, identity, free speech, and criminal liability. Any attempt to regulate deepfakes must therefore be informed by a nuanced understanding of both their technical foundations and their potential for harm.

[7] *K.S. Puttaswamy v. Union of India* (2017) 10 SCC 1.
[8] *R. Rajagopal v. State of Tamil Nadu* (1994) 6 SCC 632.
[9] Mirsky Y and Lee W, 'The Creation and Detection of Deepfakes: A Survey' (2021) 54 *ACM Computing Surveys* 1.
[10] Vaccari C and Chadwick A, 'Deepfakes and Disinformation' (2020) 16 *Social Media + Society* 1.

3. Deepfake Misuse in India

3.1 Political Misinformation and Electoral Manipulation

One of the most significant uses of deepfake technology in India has been in the political domain. During the 2024 general elections, AI-generated content was widely used to influence voter perception. Deepfake videos featuring political leaders were circulated to convey fabricated messages, often in regional languages to maximise outreach.[11]

Such misuse undermines the integrity of democratic processes by distorting public discourse. For instance, manipulated videos showing political leaders engaging in actions or making statements they never actually did can influence voter behaviour and erode trust in institutions. The use of deepfakes in political campaigning has also been commercialised, with companies offering AI-generated campaign content at scale.

3.2 Deepfake Abuse Against Women

Nonconsensual sexual content generation is a particularly serious form of misuse in India. According to surveys, 90 to 95 percent of all internet deepfakes are nonconsensual sexual images, with around 90 percent showing women. In 2023, the total number of deepfake videos online climbed by 550 percent from 2019. Deepfake erotica makes up 98 percent of all deepfake videos on the internet, and 99 percent of those targeted are women. AI enables the creation of interactive deepfakes that impersonate humans and start online conversations with women and girls who have no idea they are speaking with a bot. The practice of "catfishing" on dating sites can now be scaled out and made more realistic as AI bots learn to replicate human interactions, luring women and girls into exposing private information or meeting up offline. Victims frequently experience severe psychological distress, reputational damage, and societal humiliation.

Reports highlight that women in India are increasingly withdrawing from online platforms due to fear of being targeted by deepfake tools, including “nudify” applications.[12] Such misuse constitutes a violation of privacy, dignity, and bodily autonomy, raising serious constitutional concerns.

[11] World Economic Forum, *Deepfakes and Indian Elections* (2024). (World Economic Forum)
[12] The Guardian, *Impact of Deepfakes on Women in India* (2025).

3.3 Financial Fraud and Identity Theft

Deepfake technology has also been used to perpetrate financial fraud. In India, there have been instances where deepfake videos impersonated financial experts or corporate executives to promote fraudulent investment schemes.[13] These scams take advantage of people's faith in public persons and cause enormous financial losses.

Furthermore, voice cloning technology have allowed scammers to imitate others and trick victims into sending money. This type of cybercrime marks a new frontier in fraud, making existing authentication systems worthless.

3.4 Celebrity and Personality Rights Violations

Celebrities have frequently been the victims of deepfake usage. Courts in India have begun to recognise the gravity of such crimes. In a landmark case, the Bombay High Court ordered the removal of deepfake content featuring a well-known actor, acknowledging the violation of personality rights and the risk of public harm.

Such cases underscore the importance of legal protection of personality rights and stronger enforcement tools to protect individuals from illegal exploitation of their likeness.

4. Indian Judicial Approach

Deepfake technology has compelled courts to address new forms of digital injury that contradict existing legal principles. While there is no comprehensive regulatory framework in India to address deepfakes, the judiciary has begun to respond by modifying existing principles of constitutional law, tort law, and criminal jurisprudence.

The discussion illustrates that, despite statutory limitations, Indian courts have used established jurisprudence to address the repercussions of synthetic media. Simultaneously, comparative judicial techniques show the potential and limitations of existing legal frameworks. Indian courts have begun addressing deepfake-related issues, albeit within the framework of existing laws. In cases concerning personality rights, courts have recognized the need of protecting individuals against illegal exploitation of their likeness. The Bombay High Court's intervention

[13] Reuters, *Deepfake Investment Scam Warning by NSE* (2024).

in cases involving deepfake usage reveals judicial concern for the matter. However, such choices are frequently reactive and lack a clear legal foundation.

India does not have a law specifically addressing deepfakes. As a result, judges rely on fragmented provisions from the Information Technology Act[14], criminal law, and intellectual property law.

Proving the authenticity or untruth of deepfake content presents considerable issues. The sophistication of AI-generated media hampers forensic investigation and calls into question the admissibility of evidence. Deepfakes are frequently generated and shared across borders, making enforcement challenging. Jurisdictional limits impede effective legal action against criminals.

4.1 Constitutional Jurisprudence: Privacy, Dignity, and Identity

Article 21 of the Indian Constitution preserves the right to life and personal liberty, ensuring strong protection from deepfake damages. In *Maneka Gandhi v. Union of India*[15], the Supreme Court enlarged the definition of "life" to encompass the right to live with dignity. This interpretation expanded Article 21 to encompass substantive rights like as privacy, reputation, and autonomy. Deepfakes that degrade, sexualise, or mislead individuals violate human dignity. Deepfakes remove persons' agency and reduce them to computational objects.

In *Anuradha Bhasin v. Union of India*[16], the Court ruled that freedom of speech and expression includes the internet. However, this independence is not absolute. The state has a reciprocal duty to safeguard citizens from fraudulent and damaging communication. Article 19(1)(a) does not provide constitutional protection for synthetic falsehood causing reputational harm.

The verdict in *K.S. Puttaswamy v. Union of India* marks a watershed moment in Indian constitutional law. The Supreme Court clearly identified the right to privacy as a basic right under Article 21, which includes informational privacy, bodily autonomy, and decisional freedom. From a doctrinal perspective, deepfake usage directly implicates informational self-determination, a subject highlighted by Puttaswamy. The Court ruled that individuals must

[14] The Information Technology Act, 2000.
[15] Maneka Gandhi v. Union of India, AIR 1978 SC 597
[16] Anuradha Bhasin v. Union of India (2020) 3 SCC 637.

have control over the broadcast and use of their personal information. Deepfakes undermine control by

- extracting biometric identifiers from facial characteristics and voice patterns.
- Converting them into fabricated content.
- Disseminating this work without consent.

Furthermore, the judgment's emphasis on human dignity establishes a solid moral foundation for addressing deepfake damages. Non-consensual deepfake pornography, for example, is a serious breach of human dignity, autonomy, and sexual privacy. The Court's definition of privacy as basic to dignity enables such harms to be viewed as constitutional wrongs rather than legislative infractions.

In *R. Rajagopal v. State of Tamil Nadu*, the Supreme Court defined the right to privacy as the "right to be left alone," ruling that unauthorized publishing of a person's private life violates Article 21. This principle becomes more important in the setting of deepfakes. Deepfakes, unlike typical media publications, are fabricated rather than disclosed. The harm is thus twofold: violation of privacy due to unauthorized use of identity and creation of misleading myths that harm reputation. The Rajagopal decision tacitly indicates that even genuine publication might be limited in some circumstances; as a result, false and harmful content generated by deepfakes requires closer inspection.

Indian courts have gradually accepted personality rights, notably in situations involving celebrities and popular figures. These rights include control over one's name, image, likeness, and voice, and are based on the larger principles of privacy and property. In recent cases involving illegal digital content, courts have issued injunctions to prevent identity theft. Notably, in *Anil Kapoor v. Simply Life India*[17], the Delhi High Court recognised the actor's personality rights and prohibited the illegal use of his image and voice via digital alteration.

5. Criminal Liability and Limitations of Existing Framework

Sections 66E (violation of privacy) and 67 and 67A (obscene content) of the IT Act can be

[17] *Anil Kapoor v. Simply Life India* (2023) Delhi High Court.

used to address deepfake usage. However, these regulations are not explicitly intended to combat deepfake technology and frequently fail to capture its entire reach.

The Bharatiya Nyaya Sanhita modernises criminal law by incorporating "visible representations" into the definition of defamation. This expansion allows courts to handle reputational injury from synthetic videos and altered media. Cheating, defamation, impersonation, and forgery are among potential offences associated with deepfake usage. However, because of the novelty of the technology, the implementation of these provisions is frequently insufficient.

The DPDPA[18] requires specific consent to process personal data. Unauthorised use of biometric identifiers in AI training pipelines violates statutory duties, as they are considered personal data. However, enforcement methods are still inadequate.

Existing laws are reactive rather than preventive. They do not address issues such as:

- Creation of synthetic identities
- Manipulation of biometric data
- Platform responsibility

There is a need to recognise deepfake misuse as a distinct category of cybercrime. Criminal liability should extend to the creators as well as distributors of malicious deepfakes. Action also needs to be taken against platforms that fail to act despite knowledge of such content being circulated.

6. Legal and Regulatory Reforms

Regulating deepfake technology entails balancing innovation and fundamental rights. Given the magnitude, speed, and severity of the harm inflicted by synthetic media, gradual or reactive legal restrictions are insufficient. Instead, a complete, rights-centric framework must be implemented—one that recognizes identity as a legally protected interest in the digital domain and embeds accountability throughout the AI lifecycle.

[18] The Digital Personal Data Protection (DPDP) Act, 2023.

Firstly, Mandatory digital provenance and watermarking requirements must be institutionalized. All generative AI systems that can create realistic audio-visual content should be legally obligated to include tamper-resistant, machine-readable watermarks in their outputs. Watermarks should provide metadata about the tool used, time of production, and if biometric data was included. Watermarking duties should be implemented at the developer and service-provider levels rather than on individual users to ensure overall compliance. A framework would allow for rapid detection, attribution, and verification of generative technologies without prohibiting them entirely.

Second, the law must explicitly recognize the "Right to Digital Integrity" as an extension of fundamental personality rights. This right should give individuals enforceable control over the creation, alteration, and distribution of their digital likeness, which includes their face, voice, and behavioral characteristics. Digital integrity, unlike traditional privacy rights, aims to prevent unauthorized emulation, not only data exploitation. Codifying this right, whether through independent legislation or modifications to the Digital Personal Data Protection Act, would enable explicit civil remedies such as injunctions and statutory damages, harmonizing with dignity jurisprudence under Article 21.

To address the urgent nature of deepfake harm, judicial and administrative remedies must be expedited. Conventional litigation is not effective in dealing with synthetic media, which may cause reputational damage within hours. Creating specialized cyber tribunals or deepfake reaction benches with the ability to issue dynamic injunctions (orders that apply across platforms and search engines) could greatly minimize victim hardship. These bodies should have the authority to direct intermediaries to preserve evidence, divulge origin data, and cooperate with forensic investigations.

Fourth, intermediary accountability should move from passive compliance to active risk management. Although India's IT Rules, 2021 require takedowns, they do not sufficiently promote proactive detection. Platforms hosting user-generated content should implement AI-based deepfake detection systems and provide regular Deepfake Transparency Reports. Reports should include information on detected synthetic material, reaction timings, error rates, and prevention strategies. Transparency obligations promote public trust and regulatory monitoring without excessive restriction.

Finally, policies must include safeguards against misuse and overreach. Deepfake regulations

should contain exclusions for satire, parody, research, and genuine artistic expression, with clear labelling and no malevolent intent. To prevent surveillance misuse, law enforcement must adhere to rigorous data protection requirements when accessing biometric and provenance data. To effectively manage deepfakes, it's necessary to prioritize identity protection over content management. India can create a regulatory architecture that prioritizes dignity, democratic debate, and ethical AI governance by combining technological safeguards with constitutional values and procedural efficiency.

7. Conclusion

The harmful application of deepfake technology poses a huge challenge to judicial systems globally. In India, the rapid development of deepfakes has highlighted the shortcomings of existing legal frameworks in dealing with developing forms of digital harm. Deepfake usage has far-reaching and diverse implications, including political misinformation, gendered abuse, financial fraud, and identity theft.

While Indian courts have begun to acknowledge the gravity of the situation, the lack of a specialised legal framework restricts the efficacy of judicial interventions. Existing regulations, such as the Information Technology Act and penal prohibitions, only offer partial relief and fail to address the distinct characteristics of deepfake technology.

This study emphasized the critical need for significant regulatory reforms. A specialized deepfake regulation, together with increased platform accountability and technological controls, is required to limit the threats posed by this technology. Furthermore, improving privacy and personality rights, as well as increasing digital knowledge, can help alleviate the issue.

Finally, regulating deepfakes demands a balanced approach that protects individual rights while encouraging innovation. As technology advances, legal frameworks must change to guarantee that the benefits of artificial intelligence are not outweighed by the potential for harm.