Jus Corpus Law Journal
Open Access Law Journal – Copyright © 2024 – ISSN 2582-7820
Editor-in-Chief – Prof. (Dr.) Rhishikesh Dave; Publisher – Ayush Pandey

Tackling the Multifaceted Legal Dilemmas of Deep Fake Technology

Vaishnavi Kulkarni[a] Bhanusshre Sivaramachandran[b]

[a]Ramaiah College of Law, Bangalore, India [b]Symbiosis Law School, Hyderabad, India

Received 26 February 2024; *Accepted* 27 March 2024; *Published* 30 March 2024

Deepfake technology lies at the heart of artificial intelligence, rapidly expanding its presence across numerous social media platforms. This swift advancement urges us to delve into the legal dimensions concerning deepfakes. Exploring the implications of deepfakes in defamation, intellectual property law, and cyber law illuminates the diverse legal entanglements associated with this technology. Examining the connection of defamation with deepfakes underscores the threat of deepfakes in misinformation especially in politics and defaming celebrities. Further delving into the relation of deepfakes with intellectual property rights enables the understanding of complex copyright issues wrangled within AI-generated content. This article also emphasizes the deeply interconnected concepts of cyber law and deepfake shedding light on the concern of deepfake pornography, sextortion, fraud, and social engineering attacks. Moreover, this research imparts knowledge of the various legal provisions currently involved in addressing the deepfake issue in India. It further explores the Centre's plans to address this in the upcoming era of Digital India and provides suggestions and recommendations to enrich the regulatory framework drawing inspiration from the regulations implemented worldwide.

Keywords: *deep fake, defamation, copyright, cyber law, privacy, regulation, ai.*

INTRODUCTION

As Artificial Intelligence continues to advance and multiple technologies emerge, the prevalence of deepfake videos is steadily increasing. People are reaping the rewards of contemporary technology.

The term 'deepfake' was originally coined by a Reddit user who used 'Face-swapping' technology to superimpose celebrities' faces on pornographic videos.[1] These synthetically modified images and videos are developed through Generative Adversarial Networks.[2] A generative adversarial network system comprises two deep neural networks—the generator network and the discriminator network. Both networks train in an adversarial game, where one tries to generate new data, and the other attempts to predict if the output is fake or real data.[3] The generator attempts to maximize the probability of a mistake by the discriminator, but the discriminator attempts to minimize the probability of error. In performing this, they reach a state of equilibrium where it is no longer possible to recognize that it is synthetic data.

Although deepfakes have a wide scope for creative applications in the entertainment industry aimed at enhancing and recreating scenes and visuals, there are two sides to a coin. Recently, heightened ethical and legal concerns have arisen due to the growing misuse of this technology. The striking accuracy of deepfake videos makes them highly deceptive, requiring greater scrutiny to distinguish whether it is real or not.

In India, there is a lack of an explicit legal framework to handle deepfake technology. However, certain existing laws address those individuals who misuse this technology, holding them accountable. The use of deepfakes has resulted in increasing cases relating to defamation, copyright, and cybercrimes.

[1] Hannah Smith and Katherine Mansted, *Weaponised deep fakes: National security and democracy* (Australian Strategic Policy Institute 2020)

[2] Todd C. Helmus, *Artificial Intelligence, Deepfakes, and Disinformation: A Primer* (RAND Corporation 2022)

[3] 'What are some use cases of generative adversarial networks?' (*Amazon Web Services*) <https://aws.amazon.com/what-is/gan/#:~:text=A%20generative%20adversarial%20network%20system,is%20fake%20or%20real%20data> accessed 23 February 2024

DEFAMATION AND DEEPFAKES - A FABRICATED REALITY

With the present cutting-edge technology, the intersection between deep fake technology and defamation highlights a significant concern since manipulated videos and audio can falsely depict individuals involved in defamatory acts or statements. It poses a threat to reputation and the well-being of individuals. Various instances have come to light about this issue, encompassing legal actions involving the masses and high-profile figures such as celebrities and politicians. For example, a recent deepfake video featuring actress Rashmika Mandanna stirred a viral storm across the internet. The video showcased the British-Indian influencer, originally dressed in black workout attire, with her face seamlessly replaced by that of the Bollywood actor.[4]

In India, section 500 of the Penal Code[5] provides punishment for defamation. The victim of a deep fake video has the right to approach the court for defamation if the following are satisfied:

- Establish that the video encompasses false information or misleadingly presents the person.
- Establishes that the content is disclosed to a third party or published in some manner.
- The subject experiences reputational damage and harm due to the deceptive video.
- In certain instances, the burden of proof is on the plaintiff to prove that the video's creator acted with deliberate malice or negligently.

The offender can also be charged under sections 465[6] and 469[7] of the Indian Penal Code 1860, on the subject of forgery and harming reputation, respectively.

[4] 'Legal Implications of Deepfake image like that of Rashmika Mandanna and Katrina Kaif usage in India' *The Times of India* (8 November 2023) <https://timesofindia.indiatimes.com/entertainment/hindi/bollywood/news/legal-implications-of-deepfake-image-like-that-of-rashmika-mandanna-and-katrina-kaif-usage-in-india/articleshow/105065690.cms>accessed 11 February 2024

[5] Indian Penal Code 1860, s 500

[6] Indian Penal Code 1860, s 465

[7] Indian Penal Code 1860, s 469

Deepfake in the realm of politics: The electoral system in India is currently at risk of deepfake-enabled propaganda. This 'threat to democracy', as quoted by IT Minister Ashwini Vaishnaw[8] is used to spread misinformation to change the thoughts and perspectives of voters and portray the opposing party members in a bad light. These deepfake videos are shared with the admin control of groups for a small group of audience or to social media influencers. These groups, which are formed to circulate misinformation and political propaganda, are known as 'Scratch groups', target people from the age group of 18-25, and are tailored for distribution exclusively through WhatsApp since, unlike other social media platforms, information shared cannot be reported as 'spam' or 'suspicious'.[9]

The prevalence of deepfake in politics is not only confined to India but is seen across the global landscape. In March 2022, a deepfake video of President Vlodymyr Zelensky emerged on social media, wherein he was found urging Ukrainians to lay down their arms and surrender.[10] We have reached a stage where we can no longer rely on the veracity of what we see and hear. Defaming politicians and manipulating the voters with false videos of these politicians has become the new norm in the new political era.

Defamation in the age of misrepresentation: In the contemporary marketing landscape, where influencer marketing and celebrity endorsements are heavily relied on to engage audiences, the emergence of deepfake videos adds a new level of vulnerability leading to misrepresentation of products. This threat extends beyond the potential harm to the promoted products. It not only impacts consumer trust, but also the reputations of the celebrities involved. The unauthorized use of celebrities' images is exploitative and can lead to copyright battles between the creators of deepfakes and the celebrities.

[8] Abdul Aleem Sherif, 'Deepfake Elections: How Indian Politicians Are Using AI-Manipulated Media to Malign Opponents' (*Outlook India*, 24 November 2023) <https://business.outlookindia.com/technology/deepfake-elections-how-indian-politicians-are-using-ai-manipulated-media-to-malign-opponents> accessed 15 February 2023

[9] *Ibid*

[10] Bobby Allyn, 'Deepfake Video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn' (*NPR*, 16 March 2022) <https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia> accessed 17 February 2024

INTERSECTION OF DEEPFAKES WITH INTELLECTUAL PROPERTY LAW

In the matter of Intellectual Property Rights, many trademark and copyright concerns are looming over AI-generated content. From the Authors Guild Association filing a copyright infringement suit on OpenAI for the unauthorised use of their books by ChatGPT[11] to filing trademark claims for false endorsement using deepfake technology; there are umpteen cases of AI and deepfakes entangling with IPR laws.

There are two main contentions regarding copyright and deepfakes. Firstly, the determination of the 'Author' or 'Creator' of deepfakes, and secondly, how copyright law addresses the challenges posed by this technology.

Under section 2(d)(vi)[12] of the Indian Copyright Act 1957, the author or creator of deepfake may be the person who used the technology to create the image. However, the law does not expressly address whether AI-generated content is copyrightable. A noteworthy example in this regard would be how RAGHAV (AI-based application)[13] was initially granted co-authorship for the work produced by it but this was later objected to by the Copyright Office. Other jurisdictions worldwide hold different opinions on the same; while the USCO has clearly stated that copyright will only be granted to natural persons[14] and therefore AI-generated works cannot be copyrightable, the Beijing internet court held another perspective stating that AI-generated images are considered protectable works, and the AI user is recognized as the author. [15]

[11] Tori Guidry, 'OPENAI SUED AGAIN IN A STAR-STUDDED COMPLAINT: John Grisham And George R.R. Martin Amongst Authors Suing OpenAI In A Major Class Action' (*The National Law Review*, 5 October 2023) <https://www.natlawreview.com/article/openai-sued-again-star-studded-complaint-john-grisham-and-george-rr-martin-amongst> accessed 12 February 2024

[12] Indian Copyright Act 1957, s 2(d)(vi)

[13] Shristi Ohja, 'Who owns AI-generated works?" Here's what the laws say on Copyright issue' (*India Today*, 22 September 2023) <https://www.indiatoday.in/law/story/chatgpt-ai-generated-content-copyright-ownership-complexities-india-2439165-2023-09-22> accessed 12 February 2024

[14] Tony Analla, 'Zarya of the Dawn: How AI is Changing the Landscape of Copyright Protection' (*Harvard Journal of Law and Technology*, 6 March 2023) <https://jolt.law.harvard.edu/digest/zarya-of-the-dawn-how-ai-is-changing-the-landscape-of-copyright-protection> accessed 13 February 2024

[15] Tingting Wen, 'Beijing Internet Court recognizes copyright in AI-generated image' (2024) 19(3) Journal of Intellectual Property Law & Practice <https://doi.org/10.1093/jiplp/jpad127> accessed 13 February 2024

Deepfake is often subjected to copyright violation as its work is sourced from the images and paintings of other artists. It is a well-established fact that AI produces its image by manipulating the existing images. Most of these images are copyrighted material therefore unauthorised use of it by the creators of deepfake can give rise to legal complications.

FUSION OF PERSONALITY RIGHTS AND IPR IN THE AGE OF DEEPFAKES

Many actors and famous personalities have filed a suit of copyright against these morphed images and videos claiming to be violative of their personality and publicity rights. In the case of Anil Kapoor v Simply Life India and Ors,[16] the actor Anil Kapoor asserted multiple legal claims against the sale and use of images, taglines, and deepfake videos. It was held that unauthorized usage of a celebrity's images, GIFs, and videos for commercial exploitation is violative of the actor's right to endorse products. Moreover, deepfakes can tarnish the reputation of the actor. Thereby, concluding the case in favour of the plaintiff. This sets a landmark precedent for the violation of personality rights concerning generative AI and deepfakes. Furthermore, section 57 of the Copyright Act provides exclusive rights to authors; economic rights and moral rights.[17] Deepfakes are a distortion/mutilation of the original art/picture produced by the author which may cause harm to the reputation of the author and can be subjected to copyright violation under this provision.

The claim for copyright infringement may prove to be unsuccessful in countries like the USA wherein the 'Fair use' doctrine is applicable However, in India the approach towards fair dealing is different. Section 52[18] of the Act provides an exhaustive list of what does not classify as copyright infringement and deepfake does not fall under that list making it liable to infringement. This may be frowned upon as it fails to recognize the uses of technology for entertainment, education, and other positive endeavours.[19]

[16] *Anil Kapoor v Simply Life India and Ors* CS (COMM) 652/2023

[17] Indian Copyright Act 1957, s 57

[18] Indian Copyright Act 1957, s 52

[19] Akhtar Hussaina, 'DeepFakes: A Challenge to Copyright Law' (2023) 3(4) Jus Corpus Law Journal <https://www.juscorpus.com/wp-content/uploads/2023/09/56.-Akhtar-Hussain.pdf> accessed 15 February 2024

Moreover, the proliferation of deepfake technology not only poses challenges in the realm of copyright law but also exacerbates concerns regarding cybercrimes.

PERILS OF DEEPFAKE IN THE CYBER WORLD

Deepfakes can be classified as a type of cybercrime. Cybercrime refers to a 'criminal activity (such as fraud, theft, or distribution of child pornography) committed using a computer especially to access, transmit, or manipulate data illegally'.[20] However, not all deepfakes are crimes; they are also used for entertainment purposes and in film industries. Cybercrimes through deepfakes can be divided into- pornography, identity thefts, social engineering attacks, and blackmailing.

Pornography- a gendered concern: Deepfake pornography is by far the most widely circulated type of deepfake. Unfortunately, it is seen to have affected women the most. According to a study conducted by Deeptrace, 8 out of 10 pornography websites hosted deepfake pornography.[21] This is an alarming issue as nowadays, videos can be manipulated realistically by seamlessly incorporating a person's face with explicit content. 96% of the deepfake content with obscene material targets women and objectifies them, showcasing data-driven algorithms used against women and making toxic masculinity more prevalent within the hierarchy.[22] Out of which, 99% feature female celebrities.[23]

Sexual privacy is at the top of privacy tenets and is crucial for an individual's dignity and freedom. The proliferation of deep fakes mainly raises privacy concerns and infringes on individuals' right to privacy guaranteed under Article 21 of the Constitution.[24] The biometric information, such as physical features and voice recordings of individuals, get misused to create

[20] 'Cybercrime' (*Merriam-Webster)* <https://www.merriam-webster.com/dictionary/cybercrime> accessed 17 February 2024

[21] Henry Ajder et al., 'The State of Deepfakes: Landscape, Threats, and Impact' (*DeepTrace*, September 2019) <https://regmedia.co.uk/2019/10/08/deepfake_report.pdf> accessed 19 February 2024

[22] Aarushi Anand, 'When and how will Law wake up to deepfake Technology?' *The Leaflet* (18 November 2023) <https://theleaflet.in/when-and-how-will-the-law-wake-up-to-deepfake-technology/> accessed 16 February 2024

[23] Kate Coleman, 'How Deepfakes are impacting Culture, Privacy, and Reputation' (*Status Labs*) <https://statuslabs.com/blog/what-is-a-deepfake> accessed 20 February 2024

[24] Constitution of India 1950, art 21

realistic manipulations, thereby influencing public opinion. In the landmark judgment of KS Puttuswamy & Anr v Union of India & Ors,[25] it was held that the right to privacy is an essential aspect of the right to life and personal liberty and a facet of integrity.

After a deepfake porn video of an Indian journalist was leaked, she began to receive inappropriate and offensive notifications from men, including threats of sexual violence.[26] It started when she campaigned against government corruption concerning a rape case. The situation intensified when a BJP fan account shared it, attracting an additional 40,000 shares. The influx of numerous inappropriate and offensive notifications asking for sexual favours and comments following the data breach left a lasting impact on her mental and physical well-being, which led her to the hospital.[27] One of the serious implications related to deepfake pornography is blackmailing, extorting, seeking financial gains, or controlling the victim. She was also known as sextortion. The threat of exposed deepfake videos can have overwhelming consequences on the victim concerning reputation, dignity, and relationships, and the offenders exploit them until their demands are met.

One such AI tool aimed at exploiting women is apps and websites like Deepnude. Deepnude is a computer app that is based on deepfake technology that allows one to 'strip' off the clothes of women just by their photos. These algorithms are specifically trained to be performed on women and cannot perform similar actions on men. Although the original creators have taken down the app, it is said that what is put up online can never be removed. Therefore, it was beyond the control of the creators and such software is readily available online giving easy access to creating non-consensual pornographic content.

The Indian legislative Acts such as the Information Technology Act 2000, the Indian Penal Code, the Data Personal Protection Act 2023, and the Protection of Children from Sexual Offenses Act,

[25] *KS Puttuswamy & Anr v Union of India & Ors* (2017) 10 SCC 1

[26] Anne Pechenik Gieseke, 'The New Weapon of Choice: Law's Current Inability to Properly Address Deepfake Pornography' (2020) 73(5) VANDERBILT LAW REVIEW 1479

[27] Rituparna Chatterjee, 'I couldn't talk or sleep for three days: Journalist Rana Ayyub's horrific social media ordeal over fake tweet' (*Dailyo*, 26 April 2018) <https://www.dailyo.in/variety/rana-ayyub-trolling-fake-tweet-social-media-harassment-hindutva-23733> accessed 12 February 2024

2012 addresses pornography. For instance, sections 66E, 67, 67A, and 67B of the IT Act, 2000[28] deal with punishments for violation of privacy, publishing or transmitting obscene material in electronic form, sexually explicit material, and for publishing or transmitting sexually explicit acts depicting children respectively. However, these laws do not specifically address AI content such as deepfake.

Identity theft and social engineering attacks: A recent video of Virat Kohli endorsing a betting app promoting high profits went viral on social media platforms. It was later discovered that it was a deepfake video created by cyber fraudsters. Through deepfake algorithms, it has become extremely effortless to defraud or steal someone's identity. With the technologies of voice cloning and face-swapping, one can use the identities of celebrities to misguide the audience. In addition to that, cybercriminals can use the victim's voice and face to attack biometric systems to steal account numbers or unique identification numbers. When criminals manipulate victims directly to reveal confidential information, it is known to be social engineering.

Blackmailing: Furthermore, blackmailing, particularly, sextortion has become increasingly common in this era of digitalization. A case to this point would be when a retired IPS officer from Uttar Pradesh was blackmailed with a deepfake video of him soliciting sex. Afraid of tarnishing his reputation, the poor man had repeatedly made payments to the scammers.[29]

Section 66-C of the IT Act, 2000 provides punishment for identity theft and section 66-D punishes anyone who personates with the intention of cheating using a computer device.

The Ministry of Electronics and Information Technology had issued an advisory to all intermediaries to comply with the existing rules i.e., Rule 3(1)(b) within the due diligence section of the IT rules which mandates intermediaries to remove any content that is prohibited under the law.[30] In addition to that, MEITy held 2 Digital dialogues in which the concern surrounding

[28] Information Technology Act 2000, ss 66E, 67, 67A and 67B

[29] Ravish Ranjan Shukla Edited By Chandrajit Mitra, 'Retired Top Cop Latest Victim Of Deepfake, Video Used To Con Ghaziabad Man' *NDTV* (30 November 2023) <https://www.ndtv.com/ghaziabad-news/retired-top-cop-latest-victim-of-deepfake-video-used-to-con-ghaziabad-man-4621233> accessed 22 February 2024

[30] Ministry of Electronics & IT, MeitY issues advisory to all intermediaries to comply with existing IT rules (2023)

deepfakes by the Prime Minister was addressed. Although the government has addressed the concern surrounding deepfakes by issuing notifications and statements, a robust legal framework regarding the same is yet pending.

URGENCY FOR REGULATORY GUIDELINES

There is a dire need for a regulatory framework in India. While the consensus on this is strong and is widely acknowledged, there remains a lack of clarity in decoding the specific measures for tackling the intricate issues of deepfakes and AI. It is required to transition from emphasizing the necessity to acting upon it and developing comprehensive laws. Fundamental questions surrounding it, including whether standalone legislation is required, or it can be integrated into the existing laws. An outright ban on deepfakes is not being suggested; rather, it emphasizes the need to regulate manipulated content that harms individuals and ensure a fair integration of AI in society.

This issue has not gone unnoticed by the Centre. In a meeting comprising of the Minister of Electronics and Technology and social media platforms like Google discussed that they will join hands with the government to end this problem and comply with the rules. BSA, a global association of software companies, recommended that MEITy hold back from adopting a 'one-size-fits-all' approach for the suggested policy amendments to IT Rules and focus on the distinction between the role and functions of social media intermediaries concerning the dissemination of deepfakes.[31] Proactive initiatives by large tech companies and global industries can help develop an approach that promotes technological advancement and legislative techniques. Ashwini Vaishaw expressed the government's intentions to bring out new regulations on the detection and prevention of deepfakes, strengthening reporting mechanisms, and spreading awareness.[32]

[31] Sujit Chakraborty, 'AI Deepfake: The Indian Approach' (*The Processor*, 19 January 2024) <https://theprocessor.in/sectors/ai-deepfake-indian-2392379> accessed 25 February 2024

[32] 'Centre Planning New Regulations, Penalties for Both Creators and Platforms To Deal with Deepfakes' *Deccan Herald* (23 November 2023) <https://www.deccanherald.com/india/centre-planning-new-regulations-penalties-for-both-creators-and-platforms-to-deal-with-deepfakes-2782372> accessed 25 February 2024

The MEITy is considering amending the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 mainly focusing on 3 amendments.[33]

Defining Deepfake: To bring about any regulations relating to it, it is essential to understand what falls within the ambit of the term 'Deepfake'. This is also crucial to safeguard those modifications in images, videos, and audio that are meant for the enhancement of film and entertainment industries. Therefore, the definition would be limited to those which cause harm.

Compliance with Rule 3(1)(b):[34] This has already come into effect through an advisory notice issued to the intermediaries reminding them about the type of content that is not allowed to be displayed on their platforms.

Expanding the scope of grievance: Under the IT Rules, 2021 grievance is defined as, 'any complaint, whether regarding any content, any duties of an intermediary or publisher under the Act, or other matters about the computer resource of an intermediary or publisher, as the case may be'. This does not include user-generated content, by including that, the victim may directly approach the 'grievance officer' for a faster seek of redressal.

In addition to this, the proposed Digital India Bill, 2023 will also contain regulations for AI. As far as deepfakes and synthetic content are concerned, the government is planning to address this by issuing dos and don'ts for the intermediaries. However, it is not enough that social media platforms should be held responsible for deepfake content since they always have the option of reverting to the Safety Harbour clause as per section 79 of the IT Act.[35]

RECOMMENDATIONS

[33] Aditi Agarwal, 'Centre Likely To Amend IT Rules To Define Deepakes' *Hindustan Times* (06 January 2024) <https://www.hindustantimes.com/india-news/centre-likely-to-amend-it-rules-to-define-deepfakes-101704482610756.html> accessed 25 February 2024

[34] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021, r 3(1)(b)

[35] Information Technology Act 2000, s 79

It is proposed that rules and regulations for deepfake and AI be accommodated in the existing laws requiring intensive application and implementation. A framework is required to curb harmful deepfakes before their dissemination.

- A mandate must be put on creators and providers to obtain consent from people who are being displayed in the video, ensure the authenticity of users' identities, and provide recourse mechanisms to those affected. The service providers must set up guidelines and service agreements, put in place a system to confirm users' real identities, develop a database to detect illegal and false information, and document network logs, drawing inspiration from China's regulations on deep synthesis.[36]
- There must be a strict security evaluation while offering templates and other models and resources for editing and morphing face, voice, and physiological data, which entails public interests, national security, etc.
- There must be a greater emphasis on making intermediaries liable to verify the authenticity of videos through highly adept content moderators. Provisions should be made to ensure that the intermediaries do not fall back into the safety net provided by section 79 of the IT Act.[37]
- Mirroring the strategy adopted in the EU AI Act, India can follow a risk-based approach, wherein the degree of the risk determines the kind of rules required. Therefore, the higher the risk, the stricter the rules. More stringent regulations govern high-risk AI before the video can be disseminated. It enables high-risk assessments and precautionary systems by recording activity to detect the outcomes and documentation containing all the details about the systems and their objectives to monitor their compliance.
- Utilizing blockchain technology helps distinguish between authentic and manipulated content by timestamping and recording data on the blockchain, offering clear evidence of when the content was created. This way unauthorized alterations can be identified.

[36] Giulia Interesse, 'China to ReGULATE Deep Synthesis (Deepfake) Technology Starting 2023' (*China Briefing*, 20 December 2022) <https://www.china-briefing.com/news/china-to-regulate-deep-synthesis-deep-fake-technology-starting-january-2023/> accessed 21 February 2024

[37] Centre Planning New Regulations, Penalties for Both Creators and Platforms To Deal with Deepfakes (n 32)

CONCLUSION

A comprehensive approach must be established that fosters innovative technological advancements while safeguarding individual rights and societal values and creating a legal landscape for regulating the same. As deepfakes veil the distinction between truth and deception, the evolution of regulatory frameworks is required to uphold ethical standards and legislators must address issues of privacy, intellectual property rights, defamation, and cyber security. Most importantly, striking a balance between the freedom of expression and minimizing potential harm is imperative. Navigating this is a challenge and a complex interplay that requires considering various factors. The emphasis is on promoting a fair and equitable integration of deepfake technology in society involving responsible usage and implementation in various domains.