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

Unmasking Deepfakes – A Legal Perspective

Abhinaba Datta[a] Subarno Banerjee[b]

[a]Law Graduate [b]Calcutta University, Kolkata, India

Received 26 May 2024; *Accepted* 28 June 2024; *Published* 01 July 2024

The prevalence of deepfake technology poses a significant threat to truth and credibility in India's vastly diverse society. These are created using Artificial Intelligence consisting of complex algorithms, such as Generative Adversarial Networks (GANs), which blend 'deep learning' and 'fake' to generate images and videos that are visually indistinguishable from reality. Deepfakes have been widely used in the entertainment industry which results in opening new avenues for creators by democratizing the field and making it accessible to all. The deepfake technology was initially intended for entertainment purposes, and is now being used to spread misinformation, and non-consensual pornography and manipulate the media to tackle this problem effectively, a multi-faceted approach is necessary. In India, the Information Technology Act, 2000[1] and the Indian Penal Code, 1860[2] provide legal provisions but they are insufficient. Legal bodies such as the Ministry of Electronics and IT have taken steps to address misleading content but a deeper impact requires reinforcement of the present laws. Implementation of various measures such as public awareness campaigns and promoting media literacy can be crucial to combat the effects of deepfakes. The government of India has collaborated with various tech platforms to raise awareness by using labels to identify the usage of AI-generated media in content over the internet. The article lays emphasis on the urgency of implementing comprehensive regulations such as - strengthening legal frameworks, raising public awareness, and collaborating with other nations. India, therefore by implementing these measures can mitigate the harmful effects of deepfakes and preserve trust among the public in a world where technology is ever-evolving.

[1] Information Technology Act 2000
[2] Indian Penal Code 1860

Keywords: *deepfake, artificial intelligence, legal provisions.*

INTRODUCTION

With the advent of technological advancements, the line between truth and fiction is becoming blurred. The rise in deepfake content is a significant threat to the trust and credibility of the public, particularly in India where a diverse population is facing a digital crisis due to a rapidly evolving technological landscape. The spread of misinformation in the form of manipulated media is causing significant harm to the social and political landscape of the country.

The term 'deepfake' is a combination of the words 'deep learning' and 'fake'. Deepfake AI is a tool that uses digital software or machine learning to create fake or artificial content. The content created through deepfake AI is technically false in nature but often appears convincingly real.

As defined by the Merriam-Webster dictionary, a deepfake is "*an image or recording that has been convincingly altered and manipulated to misrepresent someone as doing or saying something that was not done or said*."[3]

CREATION OF DEEPFAKES

The process involving the creation of deep fake content uses two kinds of algorithms – a generator and a discriminator, which aid in the creation and refining of fake content.[4] The output desired is sought at the beginning and a training data set is built by the generator which initially creates synthetic digital content. Meanwhile, the discriminator analyses the realistic or imitative effects in the initial version. This process aids in the creation of more realistic content which is used to spot and remove all the flaws.

[3] 'Deepfake Definition & Meaning' (*Merriam-Webster*) <https://www.merriam-webster.com/dictionary/deepfake> accessed 29 April 2024

[4] Sara H. Jodka, Manipulating reality: The intersection of Deepfakes and the law (*Reuters*, 01 February 2024) <https://www.reuters.com/legal/legalindustry/manipulating-reality-intersection-deepfakes-law-2024-02-01/> accessed 04 May 2024

The generator and discriminator algorithms combine resulting in the creation of a Generative Adversarial Network or GAN which utilizes deep learning techniques, which were developed by Ian Goodfellow and his team in 2014.[5] In the creation of deepfake AI, the system portrays the subject from different angles in the case of a photograph and in the case of a video it also deeply analyses the movement, behaviour speech patterns, etc. The discriminator runs the information multiple times to make the image or video appear more realistic. Technically, a person creating a deepfake takes up an original video and a target individual and swaps that person's face in the video resulting in such an act by the individual which he/she never did or has done.

However, most of the deepfakes created recently have been the outcome of algorithms, which are mostly aggregations of AI and Non-AI algorithms existing in various applications and software already available on the internet, which makes it effortless to use.

IMPLICATIONS OF DEEPFAKE TECHNOLOGY

The field of Artificial Intelligence, or AI, is rapidly evolving. Currently, with the introduction of Artificial General Intelligence (AGI), AI tools are becoming increasingly sophisticated and accurate in their data comprehension and analysis. The technology includes AI-generated synthetic media which has enormous potential to boost human capabilities and provide easier access to solutions.[6]

The implementation of deepfake technology holds immense potential to transform education, as it can significantly improve the engagement and interactivity of lessons. By proper utilization of deepfake technology educators can effectively animate historical figures, thereby enhancing the learning experience of students. Additionally, AI-generated content enables costly technologies such as visual effects (VFX), accessible to independent creators and authors which brings about significant benefits.

[5] Nitish Kumar, 'What is Deepfake Technology? Origin and Impact' (*Analytics Insight,* 25 June 2023) <https://www.analyticsinsight.net/latest-news/what-is-deepfake-technology-origin-and-impact> accessed 08 May 2024
[6] Jodka (n 4)

The entertainment industry is currently experiencing a significant shift as a direct result of cutting-edge technologies such as deepfake and AI-generated graphics. These breakthroughs in technology are opening exciting new avenues for creators, democratizing the field and making it more accessible for all. Moreover, they are proving to be instrumental in accelerating game development in the video gaming industry.

Advanced AI techniques, including Computer Generated Imagery (CGI) and deepfake technology, are being employed by filmmakers to create lifelike depictions of actors. A notable example of deepfake technology being used in films is the tribute paid to Paul Walker in Fast 7, where CGI and deepfake technology were utilized to create a lifelike depiction of the late actor following his untimely passing in 2013.[7]

The applications such as synthetic voices and videos extend beyond entertainment and education, with the potential to improve, accessibility, enable multilingual content creation and protect the identities of activists and journalists.

Marketing and branding applications of deepfakes are on the rise, enabling influencers and businesses to create personalized messages and reach wider audiences. In fashion retail, deepfakes enhance customer engagement by enabling virtual try-on experiences. For example, ITC partnered with AI creative firm Akool to launch the #HarDilKiFantasy campaign for its biscuit brand Sunfeast Dark Fantasy offering participants the chance to interact virtually with Bollywood stars namely, Shah Rukh Khan by sharing their screens. Recently, Zomato launched an advertisement campaign targeting specific locations starring Hrithik Roshan who is shown craving certain dishes from popular restaurants across different cities. The reality is that the actor is not actually present in the advertisement. It is created by using a Generative Adversarial Network (GAN) to identify top-rated dishes and restaurants based on the user's GPS location

[7] Carolyn Giardina, 'How "Furious 7" Brought the Late Paul Walker Back to Life' *The Hollywood Reporter* (11 December 2015) <https://www.hollywoodreporter.com/movies/movie-news/how-furious-7-brought-late-845763/> accessed 09 May 2024

from their phone. The AI is then used to learn the speech patterns and facial expressions of the actor and alters the speech based on the user's location.[8]

The intended use for deepfake AI was to create harmless and entertaining content, such as producing celebrity face swaps or inserting faces into movies. Nonetheless, as technology progressed, so did the potential for misuse. Deepfake algorithms became more sophisticated, paving the way for more realistic and seamless alterations. This raised serious concerns about possible harmful applications, such as spreading false information, damaging reputation, or influencing political outcomes.

In the year 2017, a moderator on the social media platform Reddit created a subreddit called 'deepfakes' with the purpose of showcasing videos that featured faces of famous individuals overlaid onto pre-existing pornographic content through face-swapping deepfake technology. Since then, the term 'deepfakes' has been widely used to refer to such manipulated images and videos. Unfortunately, these deepfakes are being spread extensively throughout the internet, with some prominent examples featuring public figures like Donald Trump picking fights with police officials, Pope Francis wearing a puffer jacket and Queen Elizabeth dancing while discussing the impact of technology in our society. However, it is crucial to bear in mind whether created in good faith or not, all these events are entirely fabricated and have no occurrence in the real world.[9]

One of the most widespread uses of deepfake technology is the production of non-consensual pornography, in which the bodies of pornographic actors are digitally altered to include the faces of famous or regular people. This approach infringes upon privacy and causes serious harm to the reputation of a person, as demonstrated by the 2019 arrest of an Indian male for creating a deepfake porn video that featured his girlfriend.[10] Back in the month of November of 2023, a deepfake video featuring Bollywood actor Rashmika Mandanna caused quite a stir on

[8] Sharmita Kar, 'How Advertisements Are Using Deepfake: Is There a Cause for Concern?' (*Outlook India*, 27 November 2023) <https://www.outlookindia.com/national/how-advertisements-are-using-deepfake-is-there-a-cause-for-concern--news-333087> accessed 09 May 2024

[9] 'Deepfake' (*Encyclopedia Britannica*, 08 February 2024) <https://www.britannica.com/technology/deepfake> accessed 10 May 2024

[10] Manik Tindwani, 'Deepfakes & It's Legal Implications in India' (*Law Foyer*, 27 November 2023) <https://lawfoyer.in/deepfakes-legal-implications-in-india/#google_vignette> accessed 10 May 2024

social media. The video depicted a British-Indian influencer named Zara Patel, who was wearing a black workout dress and made to falsely appear as Rashmika Mandanna while entering an elevator. This video was generated using deepfake technology, which swapped the face of Zara Patel with Rashmika Mandanna. The misleading nature of this video raised serious questions regarding cyber security, an advisory was issued to warn the public about the potential for manipulated content.[11] Similarly, with the rise in popularity of K-pop idols and actors, South Korea has seen a rise in deepfake content where numerous celebrities have fallen victim to deepfake pornography, which has caused significant emotional distress and damage to their reputation.[12]

The use of deepfakes to spread misinformation in the political realm is becoming increasingly prevalent. This involves the distribution of fabricated statements by global leaders and untrue details about political contenders. During the Delhi Legislative Assembly elections in 2020, the Bharatiya Janata Party (BJP) used deepfake technology to create AI-generated videos of its candidate, Manoj Tiwari, speaking in different languages and dialects. This was used to appeal to various linguistic groups, raising concerns about ethical practices in political campaigns.[13] In a similar case reported by Al Jazeera, a political party from southern India, Tamil Nadu, the Dravida Munnetra Kazhagam (DMK), has employed AI to recreate their iconic leader M Karunanidhi who was since deceased, showcasing realistic videos of the former screenwriter and veteran politician during their political campaign.[14] Another example of this occurred in 2018 when a video featuring a deepfake of Barrack Obama during his presidency went viral, showcasing him uttering words he actually never spoke.[15] A similar situation occurred during

---

[11] 'Who is Zara Patel, Woman in Rashmika Mandanna's Viral Deepfake Video?' *India Today* (08 November 2023) <https://www.indiatoday.in/movies/celebrities/story/who-is-zara-patel-woman-in-rashmika-mandannas-viral-deepfake-video-2459314-2023-11-07> accessed 10 May 2024

[12] Kim Bong-kee, 'K-Pop Stars Fall Victims to 'Deepfake' Porn Videos' *The Chosun Daily* (24 October 2019) <https://www.chosun.com/english/national-en/2019/10/24/5ZAK3IBVSRLZPJVYR2IRJUT4BE/> accessed 10 May 2024

[13] John Xavier, 'Deepfakes enter Indian election campaigns' *The Hindu* (03 December 2021) <https://www.thehindu.com/news/national/deepfakes-enter-indian-election-campaigns/article61628550.ece> accessed 10 May 2024

[14] Nilesh Christopher, 'How AI is resurrecting dead Indian politicians as election looms' *Al Jazeera* (12 February 2024) <https://www.aljazeera.com/economy/2024/2/12/how-ai-is-used-to-resurrect-dead-indian-politicians-as-elections-loom> accessed 10 May 2024

[15] Kaylee Fagan, 'A video that appeared to show Obama calling Trump a "dipsh-t" is warning about disturbing new trend called 'deepfakes'' *Business Insider* (18 April 2018) <https://www.businessinsider.in/a-video-that-

the height of the Russia-Ukraine conflict, a video was released that appeared to depict Ukraine's President, Volodymyr Zelenskyy, calling for his soldiers to surrender.[16] This clip sparked fury on social media, but it was later revealed to be a deepfake.

The use of deepfakes in criminal proceedings is considered a major issue in creating evidence to either provide alibis for activities or to prove the innocence or guilt of any person. This has been highlighted as a threat to the judicial system all over the world by Sara Thompson, chief product officer at BlueStar, a litigation services and technology company. The National Court Reporters Association has reported that every person is at least at a surface level, subject to such legal standards and principles that are equally enforced rather than being subjected to the personal whims of powerful corporations, individuals the government and other similar entities.[17]

Deepfakes are frequently produced to parody or satirize public figures, often depicting them in humorous or ridiculous situations. Although impersonating celebrities or politicians is widely regarded as acceptable, concerns arise when ordinary individuals are unintentionally involved in the production and circulation of deepfake content. Additionally, the widespread circulation of deepfakes for entertainment purposes could potentially lead to desensitization among the public, making it easier for malicious uses of this technology to go unnoticed.

Another concerning aspect of deepfake technology is the spread of false information on a large scale. Government and other official groups can use deepfakes to fuel tensions, causing anxiety, social turmoil and even violence. For example, a fabricated video once falsely depicted a leader of the Bahraini opposition as collaborating with Qatar to escalate hostilities, leading to further divisions. As such, it is crucial to remain vigilant and take preventive measures against these negative impacts.[18]

---

appeared-to-show-obama-calling-trump-a-dipsh-t-is-a-warning-about-a-disturbing-new-trend-called-deepfakes/articleshow/63807263.cms> accessed 10 May 2024

[16] Bobby Allyn, 'Deepfake Video of Zelenskyy Could Be "Tip of the Iceberg" in Info War, Experts Warn' (*NPR*, 17 March 2022) <https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia> accessed 10 May 2024

[17] Rachel Curry, 'AI deepfakes are poised to enter court proceedings at time of low trust in legal system' *CNBC* (06 March 2024) <https://www.cnbc.com/2024/03/06/ai-deepfakes-are-headed-to-court-at-time-of-low-trust-in-legal-system.html> accessed 27 June 2024

[18] Tindwani (n 10)

Deepfakes are a growing concern. Malicious deepfakes can lead to social manipulation, election interference, revenge porn and impersonation. To minimize the associated damages, legal frameworks that incorporate the principles of legality, proportionality and consent need to be established.

LEGAL PROVISIONS IN INDIA

The Ministry of Electronics & IT of the Union Government of India, on 7th of November 2023 issued an advisory to social media intermediaries regarding the identification of misleading information and deepfake.[19]

The key points of the advisory to the significant social media intermediaries are:

- It is imperative to apply due diligence and make reasonable efforts to uncover deepfakes and misinformation particularly that information that contravenes rules, regulations, and terms of user agreement.
- Within the periods specified by the IT Rules 2021[20], such issues are to be promptly addressed
- It causes users to refrain from hosting such data, material, or deep fakes
- When such content is reported, remove it within 36 hours of the report
- Ascertain prompt action, well within the deadlines established in the IT Rules 2021[21], and block access to such data or content[22].

The relevant provisions of the Information Technology Act, 2000[23] and rules are to be strictly ensured by the intermediaries, failure of which would attract Rule 7 of the IT Rules[24] (Intermediary Guidelines and Digital Media Ethics), 2021 and Section 79(1)[25] of the Information

[19] Ministry of Electronics & IT, 'Union Government issues advisory to social media intermediaries to identify misinformation and deepfakes: Remove any such content when reported within 36 hours of reporting' (*PIB*, 07 November 2023) <https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1975445> accessed 18 May 2024
[20] Information Technology Rules 2021
[21] *Ibid*
[22] Ministry of Electronics & IT (n 19)
[23] Information Technology Act 2000
[24] Information Technology Rules 2021, s 7
[25] Information Technology Act 2000, s 79(1)

Technology Act, 2000 would also come in action depriving the protection available to the organisation in case of any such failure.[26] It was also advised to file an FIR at the nearest police station if anyone is affected by deepfakes of themselves to seek remedies under the Information Technology Rules, 2021[27].

Although there is no such law in India that directly addresses the deepfake technology, however, there are provisions that regulate its misuse. Creating deepfake content and misusing such for breaches in cyber security, violating privacy, fraud, defamation, and copyright infringement usher legal consequences for those responsible. Individuals working with deepfake must use it responsibly, without any infringement of the rights and privacy of others ultimately preventing its legal consequences.

Apart from Rule 7 of the IT Rules Code, 2021[28] and Section 79(1)[29] of the Information Technology Act, 2000 several other sections are regulating the media content. The ministry also stated that any content that is prohibited by the IT Rules, specifically those enumerated in Rule 3(1)(b)[30], must be made explicit to users clearly and concisely, including through user agreements and terms of service. It further stated that the same needs to be made clear to the user when they first register and regularly, especially while logging in and when uploading or data sharing on the site.

Section 66D[31] provides punishment of up to three years and a fine which may extend to INR 1 lakh if anyone uses any computer resource or any communication device to cheat by personation. This was also seen in a recent case of popular actor Rashmika Mandanna where a deepfake video of her featured across social media, this section was invoked.

Section 66E[32] can be said to be applicable for deepfake crimes if anyone intentionally captures, publishes, or transmits the image of a private area of any person without his or her consent,

[26] Ministry of Electronics & IT (n 19)
[27] Information Technology Rules 2021
[28] Information Technology Rules 2021, s 7
[29] Information Technology Act 2000, s 79 (1)
[30] Information Technology Rules Code 2021, s 3(1)(b)
[31] Information Technology Act 2000, s 66 D
[32] Information Technology Act 2000, s 66 E

violating the privacy of that person, the punishment for which is imprisonment up to three years or with fine not exceeding two lakh rupees, or with both.

Apart from the above provisions; Section 67[33] any electronic publication or transmission of obscene material shall be punished on conviction with imprisonment up to 3 to 5 years on subsequent conviction and a fine up to INR 10 lakhs.

Section 67A[34] electronic publication of material containing sexually explicit acts on conviction punished with imprisonment up to 5 to 7 years on subsequent conviction and a fine up to INR 10 lakhs in both cases.

Section 67B[35] electronic publication of material depicting children in a sexually explicit act shall be punished on conviction with imprisonment up to 5 to 7 years on subsequent conviction and a fine up to INR 10 lakhs in both cases.

The Indian Penal Code, 1860 (IPC)[36] also features some provisions that can be resorted to for crimes in cyberspace which include deep fakes, these sections are applied along with the Information Technology Act, 2000[37] —

Sections 292 to 294[38] regulate the circulating of obscene materials of the acts, in the case of non-consensual pornographic material, which can damage the reputation of individuals if circulated and created using deepfake technology.

Section 354C[39] states that any man who watches or 'captures an image' of a woman doing something in private, where she expects privacy or takes part in the circulation of those images shall be liable to be punished. Explanation 2 of this section particularly mentions that if the victim consents to have their images or actions recorded but does not agree to the circulation of

[33] Information Technology Act 2000, s 67
[34] Information Technology Act 2000, s 67A
[35] Information Technology Act 2000, s 67B
[36] Indian Penal Code 1860
[37] Information Technology Act 2000
[38] Indian Penal Code 1860, s 294
[39] Indian Penal Code 1860, s 354C

those images or actions then any act of circulation of those shall be considered as an offence under this section.

Section 499 to 501[40] explains criminal defamation and punishment for such, imprisonment for 2 years with a fine or both. It can also be associated with deepfakes, which can also be used as a defamatory tool and will entice this section.

Section 504[41] and 505[42] state that intentional insults and statements conducting public mischief can also be associated with deepfakes which can provoke a breach of peace in the community and upon conviction punished with imprisonment fine or both.

Section 509[43] states that whoever intends to insult the modesty of a woman, utters any word makes any gesture or intrudes upon the privacy of such woman is punished with simple imprisonment extending to 3 years and with a fine.

Along with the above-mentioned sections, there are also a few acts that come into the picture. The following sections of the acts are also implemented in certain cases:

Section 153 (A)[44] and Section 153 (B)[45]; where deepfake technology is misused in spreading communal hatred and breaching public tranquillity, such is regulated by implementing these sections and thus preventing a toxic online environment.

Sections 468[46] and 469[47], which state that forgery and where reputation is harmed by forgery, were invoked by the police registered in an FIR in the recent case of Rashmika Mandanna deepfake. Section 420[48] is also invoked along with forgery and identity theft in cyberspace.

[40] Indian Penal Code 1860, s 501
[41] Indian Penal Code 1860, s 504
[42] Indian Penal Code 1860, s 505
[43] Indian Penal Code 1860, s 509
[44] Indian Penal Code 1860, s 153A
[45] Indian Penal Code 1860, s 153B
[46] Indian Penal Code 1860, s 468
[47] Indian Penal Code 1860, s 469
[48] Indian Penal Code 1860, s 420

Apart from these, the Copyright Act of 1957[49] comes to the scenario in case any image or video having a copyright has been manipulated to generate deepfakes. Section 51[50] provides an exclusive right on any property of any person preventing its unauthorised use by anyone.

Deepfakes have a significant effect on elections and can affect the entire process by spreading false information and fostering public opinion which tends to affect the results. Using deep fake to promote enmity and difference among people in connection with an election or to incite corrupt practices attracts Sections 123[51] and 125[52] of The Representation of Peoples Act, 1951 along with the Information Technology Act, 2000 to regulate such.

The election commission on September 26, 2019, issued a 'Voluntary Code of Ethics' during the general elections to the Haryana & Maharashtra legislative assemblies and it would also have a significant effect on all future elections.[53] This code issued guidelines for the Internet & Mobile Association of India (IAMAI) and several social media platforms such as Facebook, WhatsApp, Twitter, Google, Share-chat etc. which had been presented and to be observed accordingly during the General Election to the 17th Lok Sabha 2019. The assurance for conducting free and fair elections had also been provided by the IAMAI that all the platforms will strictly adhere to the code.[54]

On 20th March 2019, this came into effect, the same day it was presented to the Commission. There were reports from ECI that 909 such cases had been taken into action by the social media platforms during the period of the election.[55]

[49] Copyright Act 1957
[50] Copyright Act 1957, s 51
[51] Representation of Peoples Act 1951, s 123
[52] Representation of Peoples Act 1951, s 125
[53] Election Commission, '"Voluntary Code of Ethics" by Social Media Platforms to be observed in the General Election to the Haryana & Maharashtra Legislative Assemblies and all future elections' (*PIB,* 26 September 2019) <https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1586297> accessed 18 May 2024
[54] *Ibid*
[55] *Ibid*

Some highlighted features of Voluntary Code of Ethics are:[56]

- Awareness campaigns are being initiated by various social media platforms to raise informational, educational and communication initiatives along with the election rules and other pertinent guidelines.
- The ECI ensures expeditious action in reported cases, and the social media platforms have created a high-priority dedicated grievance redressal for such a purpose.
- Social media platforms and ECI have collaborated to develop a 'notification system' to alert relevant platforms of potential breaches of Section 126 of the Representation of Peoples Act, 1951 and other electoral laws and ensure strict action by the ECI.
- The media Certification and Monitoring Committees will have to make sure that every political advertisement displayed has been pre-certified by them and are in accordance with the directives of the Supreme Court.
- All the participating platforms should promote transparency in paid political advertisements, which includes harnessing their pre-existing labels/disclosure technology for such advertisements.

INTERNATIONAL SCENARIO

Several countries have introduced legislation specifically targeting deepfakes and preventing their use for any malicious aspirations. In the United States in December 2018, Senator Ben Sasse introduced the US Malicious Deep Fake Prohibition Act, 2018[57] which aims to eliminate the misuse of deepfake to defraud, extort, harass, or harm the reputation of anyone and make it a criminal offence.[58] The Deepfakes Accountability Act, 2019 introduced by US Congresswoman Yvette Clarke requires defending and combatting manipulated media and the spread of

[56] *Ibid*

[57] Malicious Deep Fake Prohibition Act 2018

[58] Felix Juefei-Xu et. al., 'Countering Malicious DeepFakes: Survey, Battleground, and Horizon' (2022) 130(7) International Journal of Computer Vision <http://dx.doi.org/10.1007/s11263-022-01606-8> accessed 18 May 2024

misinformation.[59] The bill creates guidelines on the use of this technology and imposes penalties on infractions.[60]

In following a very recent case where sexually explicit deepfake images of popular singer Taylor Swift[61] the US Government introduced the Disrupt Explicit Forged Images and Non-Consensual Edits (DEFIANCE) Act[62] on January 2024, which ensures compensation to the victims of AI-generated porn and deepfakes. In addition, on the 10th of January, US lawmakers introduced the 'No Artificial Intelligence Fake Replicas and Unauthorised Duplications (No AI FRAUD) Act,' which aims to prevent the use of personal information, especially their face and voice to create AI fakes. The terms, likeness and voice have been specifically highlighted in the bill which focuses on preventing the non-consensual use of such in creating deceiving digital content.[63]

Bletchley Declaration on AI Safety: On the Month of November 2023, a safety summit was held in Bletchley Park, United Kingdom, which focused on the risk analysis and mitigation of AI through international action.[64] This summit was joined by twenty-nine countries which includes India, the US, Germany, China, Canada, Australia, France, Japan and the European Union and many other countries with the paramount aim of assessing the opportunities and identifying the threats caused by the increasing use of AI.[65]

This Summit had some key takeaways:[66]

- The transformative potential of AI that coincides with the significant risks of AI;
- There is a desperate need to guarantee the safety of frontier AI;

---

[59] *Ibid*

[60] The Deepfake Accountability Act 2019

[61] Sarasvati NT, 'Here's a Look at US Govt's New Legislations to Tackle Deepfakes' *MediaNama* (20 February 2024) <https://www.medianama.com/2024/02/223-us-govt-deepfake-legislation/> accessed 18 May 2024

[62] Disrupt Explicit Forged Images and Non-Consensual Edits Act 2024

[63] No Artificial Intelligence Fake Replicas and Unauthorised Duplications Bill 2024

[64] The Bletchley Declaration 2023

[65] *Ibid*

[66] Steven Farmer and Johanna Lipponen, 'Key Takeaways from the UK's AI Summit: The Bletchley Declaration' (*Internet & Social Media Law Blog*, 07 November 2023) <https://www.internetandtechnologylaw.com/ai-summit-bletchley-declaration/> accessed 18 May 2024

- The most important need is to attain international cooperation which can only be achieved through an open global discussion on AI.

The main agenda of this declaration is to address the risk of frontier AI which focuses on the; Identification of AI safety risks as a shared concern building a scientific and evidence and evidence-based perception of these potent risks, and building risk-based policies like evaluation matrices and safety testing tools, development of public sector capability and research to ensure safety considering the identified and unidentified risks of AI and cooperating as appropriate while recognising the country-specific differences in approach.

The Government of the United Kingdom has taken an approach to unveil nationwide guidelines, particularly for the AI segment alongside assessing the possibility of passing legislation explicitly for the identification of images and videos/films generated by AI and controlling such content.[67]

The Digital Services Act[68] was enforced by the European Union primarily to improve transparency and aid users in discerning the credibility of any media content and the social media sites are required to comply with such necessary regulations. A law was passed by the South Korean Government which forbids distribution of deepfakes especially those endangering public safety. Any infringement of such law results in an immediate fine of up to 50 million or five years in prison.[69]

Recent reports show a large volume of misinformation is circulating across media platforms about the Israel-Gaza conflict from various accounts all over the world. This shows how political issues are misrepresented and misinterpreted which directly or indirectly affects the global scenario and yet they continue to spread endlessly.

[67] Vikrant Rana et. al., '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> accessed 18 May 2024
[68] Digital Services Act 2022
[69] Rana (n 67)

NATIONAL SCENARIO

A political scientist at the Centre for Policy Research and Visiting Assistant Professor in Political Science at Amherst College, Gilles Verniers, said that the key component of political strategy and political practice is the development of misinformation in India. The number of rumours that circulate via 'WhatsApp' in India is alarming and has led to serious issues. The broad use of messaging apps has always been a major part of the problem not only in India but also across the world.[70]

In the present scenario in India, there is no legislative framework regarding regulation and control of deepfake technology. Thus, such crimes fall beyond the purview of any one legislation. There was a proposed rule released by the Ministry of Electronics and Information Technology titled 'The Information Technology (Intermediaries Guidelines (Amendment) Rules) 2018[71], which would require intermediaries to delete or prevent access to illegal information within 24 hours of receiving a complaint'.[72] This eliminates any deepfake content that is prohibited in India. Nevertheless, India's deepfake regulations are insufficient and need to be strengthened by implementing proper and separate legislation.

The Government of India had taken an approach towards AI in the 2018, NITI Aayog's National Strategy for Artificial Intelligence. The main aim of such was to discuss the application of AI in the fields of healthcare, agriculture, infrastructure, education, transportation, and smart cities. This initiative ensured a roadmap towards research and development of AI with consideration to fairness, accountability, and transparency. A Responsible AI document was released by NITI Aayog in 2021 mainly focusing on Facial Recognition Technology and mitigating other dangers of AI. However, this does not directly address trivial topics such as deepfake but it only surfaces certain privacy-oriented issues.[73]The technology that was available at the time of its conception influenced the rules that were crafted at that time.

---

[70] Mitali Mukherjee, 'AI deepfakes, bad laws- and a big fat Indian Election' (*Reuters Institute,* 19 March 2024) <https://reutersinstitute.politics.ox.ac.uk/news/ai-deepfakes-bad-laws-and-big-fat-indian-election> Last accessed 18 May 2024

[71] Information Technology (Intermediaries Guidelines Rules) 2018

[72] Aditya Mehrotra, 'Dissecting the Framework of Deep Fakes in India – A Glaring Lacuna' (*Cell for Law and Technology*, 06 January 2024) <https://clt.nliu.ac.in/?p=887> accessed 18 May 2024

[73] Tanya Aggarwal, 'Navigating the deepfake dilemma: Government oversight in the age of AI' (*Observer Research Foundation*, 28 March 2024) <https://www.orfonline.org/expert-speak/navigating-the-deepfake-dilemma-government-oversight-in-the-age-of-ai> accessed on 18 May 2024

To make laws several stages are involved which include regulatory and parliamentary approvals, stakeholder consultation etc. and is a time-consuming process. A legal vacuum is thus, created between the development of technology and the protection of its users. A milestone measure was taken in 2023 by the passing of The Digital Personal Data Protection Act[74] (DPDP). However, this did not directly address the deepfake rather focused on the processing of individual data by companies and third-party intermediaries. The recent scenario as discussed involves provisions of the Information Technology Act 2000 and 2021[75], in addition to the Indian Penal Code[76] or the recently introduced Bharatiya Nyaya Sanhita (BNS)[77], to punish or penalise those misusing this technology.[78]

To change the current scenario, the earliest need is to alter the laws regulating deepfake to deal with the perps. In a recent example, an Executive Order 'on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence' has been signed by US President Joe Biden which labels such altered content and ensures immediate action.[79]

In the case of *Justice KS Puttaswamy (Retd.) v Union of India*[80] right to privacy was recognized as a fundamental right under Article 21[81] and any picture or video of any person cannot be used without their consent. It ensures the protection of human dignity, autonomy over information and control over one's data. Deepfakes however, violate such by breaching one's data and using them for misleading and wicked purposes.

There have also been some recent scenarios involving deepfakes that steered lawsuits in India.

In *Anil Kapoor v Simply Life India and Ors*.[82], where any deepfake content has been created using AI, using the face and voice of the Bollywood actor and several GIFs, emojis, and ringtones were created and circulated across social media platforms. A suit was filed, and the actor's attributes

[74] Digital Personal Data Protection Act 2023
[75] Information Technology Act 2000
[76] Indian Penal Code 1860
[77] Bharatiya Nyaya Sanhita 2024
[78] Aggarwal (n 73)
[79] *Ibid*
[80] *Justice KS Puttaswamy (Retd.) v Union of India* (2017) 10 SCC 1
[81] Constitution of India 1950, art 21
[82] *Anil Kapoor v Simply Life India and Ors* CS (COMM) 652/2023

and persona were protected by the Delhi High Court against any misuse for commercial gain especially in creating AI-generated deepfake content. Sixteen such persons were restrained and an injunction was granted by the court from using the actor's name in creating deepfake content.

In *Amitabh Bachchan v Rajat Negi and Ors*[83] a fake Kaun Banega Crorepati lottery fraud took place and the actor sought protection for his publicity and personality rights. An ad-interim ex-parte injunction was granted in his favour and the defendants were restrained, also any infringement of personality rights exclusive to him for commercial use is to be protected.

The latest victim of deepfake technology is Bollywood superstar Akshay Kumar, which again raised serious concerns about identity theft.[84] He expressed intense distress and aggravation after a fake video of him praising a game app went viral online last year. Recently another AI-generated video of him went viral on X (formerly Twitter), which depicts him urging users to download a so-called gaming application using statements like, 'Do you like to play too? I suggest that you download this application and attempt the Aviator game. Here, everyone plays slot machines that are most popular worldwide. Our opponents in this game are other players, not the casino'.[85] This has raised quite a concern and the actor has been prompted to file a cyber complaint. This shows that deepfake is quite a growing concern in recent times and the content though seems realistic is yet fabricated leading to the compromise of identities, especially of public figures.

In a report by India Today on the 8th of January, 2024 the High Court of Delhi gave some more time to the centre to reply to a plea addressing the issues relating to deepfake and artificial intelligence (AI) and commissioning restrictions for the same. A bench led by Chief Justice Manmohan (Acting) has given the centre a further two weeks to file its response before the court. Aware of the wider consequences of the issue, the court underlined that the response is to be provided to the relevant ministry as soon as possible. The petitioner urged the court to direct

---

[83] *Amitabh Bachchan v Rajat Negi and Ors* (2022) SCC Online Del 4110

[84] 'Akshay Kumar Takes Legal Action against Deepfake Misuse: Details Here' *The Economic Times* (03 February 2024) <https://economictimes.indiatimes.com/magazines/panache/akshay-kumar-takes-legal-action-against-deepfake-misuse-details-here/articleshow/107382839.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst> accessed 18 May 2024

[85] *Ibid*

the Centre to identify and ban such websites that provide access to deepfakes and artificial intelligence (AI), to establish rules and regulations of deepfake media content. The petitioner also urged that for both AI and deepfake access the existing guidelines are imperative and there must be an alternative set of guidelines that ensure fair implementation of AI in alignment with the Fundamental Rights. The petition also addressed the malignant adoption of AI for misinformation especially through deepfakes, citing examples of deliberate propaganda.[86]

The Delhi High Court in a recent case stated that in the age of deepfake technology, photos used to accuse someone of adultery must be backed up with solid evidence in family court. This came up when a man claimed his estranged wife was committing adultery. The court rejected his appeal paying a sum of 75,000 rupees per month to his wife and child. The court highlighted the challenge of identifying individuals in photographs due to deepfakes and noted that the husband had not mentioned an adultery claim in his response to his wife's plea. Since the divorce petition is pending, the court has allowed both parties to present evidence on the issue.[87]

In the recent advent of the nationwide election in India, there were several instances of misuse of deepfake, from AI-generated campaign videos to personalised audio messages in different Indian languages and even pre-recorded automated calls in the voice of the candidates. In a report by the BBC, News India mentions several such cases. During the ongoing campaigns, two videos went viral which portrayed Bollywood actors Aamir Khan and Ranveer Singh campaigning for Congress Party. Later it was revealed that both were created using deepfakes without their consent, and both filed a police complaint.[88]

Prime Minister Modi, on the 29th of April, raised concerns about the use of AI, for malicious and distortion of speeches by Senior Leaders of the party, including him.[89] The main issue is that due

[86] Srishti Ojha, 'High Court Grants Centre Time to Respond to Plea on AI, Deepfake Regulations' *India Today* (08 January 2024) <https://www.indiatoday.in/law/story/artificial-intelligence-deepfake-row-delhi-high-court-grants-centre-time-to-respond-to-plea-on-regulations-2486075-2024-01-08> accessed 18 May 2024
[87] 'In deepfake era, photos claiming adultery by spouse to be proved by evidence: Delhi HC' *The Economic Times* (08 June 2024) <https://economictimes.indiatimes.com/news/india/in-deepfake-era-photos-claiming-adultery-by-spouse-to-be-proved-by-evidence-delhi-hc/articleshow/110823310.cms?from=mdr> accessed 18 May 2024
[88] Meryl Sebastian, 'AI and deepfakes blur reality in India elections' *BBC* (16 May 2024) <https://www.bbc.com/news/world-asia-india-68918330> last accessed 18 May 2024
[89] *Ibid*

to the lack of comprehensive regulations, there are no severe action is taking place despite arrests.[90] The head of Internet Freedom Foundation in India, Prateek Waghre, alluded that the core issue is the need for a long-term measure to address the problem of consensual and non-consensual imagery especially created by AI or deepfake technology.[91] He also says that due to a lack of proper grievance mechanisms vulnerable groups have been subject to revenge porn and morphed imagery.[92]

REMEDIES TO COMBAT DEEPFAKES

The legal framework of India requires a substantial revamp to effectively tackle the subject of deepfake technology. Unlike several European nations and the United States, India currently lacks a comprehensive set of laws that are specifically designed to counter the threat posed by deepfake content. The existing legislation on defamation, privacy and cybercrime falls short of effectively addressing this intricate form of manipulation.

The necessity for stricter regulations surrounding the circulation of deepfakes has been an ongoing debate over the years which was brought to the forefront with the recent viral deepfake video featuring Bollywood actor Rashmika Mandanna. The current provisions of the information technology laws may not be effective in addressing the issue of deepfake as they cannot prevent their creation and initial spread. Policymakers must take action to address the significant harm caused by these manipulated videos and images. While the criminal provisions outlined in the Information Technology Act[93] and the Indian Penal Code (IPC)[94] offer some solutions to the harm caused by deepfakes, policymakers should focus on identifying solutions that mitigate the impacts and reduce the psychological toll on victims.[95]

The implementation of strict laws is a response to the consequences of the widespread circulation of deepfakes. However, to effectively address the issue, it is necessary to take

[90] *Ibid*
[91] Mukherjee (n 70)
[92] *Ibid*
[93] Information Technology Act 2000
[94] Indian Penal Code 1860
[95] Annapurna Roy, 'IT Act Needs Stronger Provisions to Curb Deepfake Menace: Experts' *The Economic Times* (14 November 2023) <https://economictimes.indiatimes.com/tech/technology/it-act-needs-stronger-provisions-to-curb-deepfake-menace-experts/articleshow/105190697.cms?from=mdr> accessed 13 May 2024

measures such as stronger regulation of the use of AI in social media and the implementation of public awareness campaigns.

In our society, the influence of social media is immense which requires individuals to be informed about the dangers of deepfakes to prevent their harmful impact. Therefore, it is crucial to launch comprehensive nationwide campaigns that raise awareness about the prevalence and risks of deepfakes. These campaigns should prioritize educating the public on how to identify and report suspected deepfake content. Additionally, integration of media literacy into school curricula can help younger generations with the skills to differentiate between authentic and manipulated or false content, empowering them to make informed decisions and avoid falling victim to deepfake content.

The impact of social media platforms on the circulation of deepfakes cannot be overlooked. It is necessary to enforce regulations on such practices and curb the circulation of manipulated content. Social media platforms must be held accountable for monitoring and eradicating deepfake content from their platforms. The establishment of guidelines to ban accounts and remove identified deepfakes is vital.

In 2019, Facebook, Amazon and Microsoft launched 'Deepfake Detection Challenge' to identify and develop effective technologies for combating the rise of deepfakes on the internet. However, the leading team in the competition was only able to achieve 65% accuracy in the detection of deepfakes through their system. As a result, social media platforms such as Facebook, Reddit and X (formerly known as Twitter) have established their own regulations to tackle the emergence and circulation of deepfakes in their platforms.

The social media platform X (formerly known as Twitter) has implemented a system of community notes that allows users to report false or manipulated information, doing so a note is added below such content informing users that the content is false or manipulated. YouTube has reaffirmed its commitment to address the spread of deepfakes, citing that hosting such content goes against the values of the platform. In the future, YouTube content creators will be required to disclose their utilization of Generative AI in the videos, with mandatory labelling in both the video player and description box.

In a move to tackle the potential dangers of deepfakes, the Government of India has joined forces with Google, a prominent player in the online industry. Google has taken proactive measures in addressing the issue of misleading online campaigns by utilizing both machine learning and human reviewers to detect and flag AI-generated content. The introduction of *SynthID* by Google is used to identify and detect AI-generated images by embedding inconspicuous watermarks onto such images to mark them as synthetically developed.[96] Furthermore, Google has updated its policies for election advertising, mandating publishers to disclose if their ads contain digitally altered or generated content intended to deceive users.[97]

It is worth noting that the laws and regulations surrounding AI are more comprehensive and rigorous in nations like the US and Europe. To address the issue of deepfakes, international collaboration with foreign nations can prove to be vital. India can gain valuable insights and exchange best practices with countries at the forefront of this battle. India must take an active role in international forums to establish global standards that govern the regulation of deepfake content. India by such participation can ensure that its perspectives and interests are adequately represented on a global level.

CONCLUSION

The advancement in technology comes with its pros and cons. The increase in the use of deepfakes needs systematic regulation which the current laws in India lack and therefore require a reformation to tackle this technology. No current laws in India define the term – 'deepfake' clearly. The laws against defamation, fake news, or violation of the modesty of a person in cyberspace are combined with the Information Technology Act, Indian Penal Code and some other statutes and the individual cases are tackled by the police. However, it does not provide any solution to the problem.

The internet with its vast nature makes it difficult to administer the rights of privacy as crimes are not limited by any national law. It can be difficult to prove in a court of law that a particular

[96] Curry (n 17)
[97] Sujit Chakraborty, 'AI Deepfake: The Indian Approach' (*The Processor*, 19 January 2024) <https://theprocessor.in/sectors/ai-deepfake-indian-2392379> accessed 13 May 2024

video or an audio clip is the result of deepfake technology created to spread false information or propaganda especially when technology is increasingly getting complicated and the credibility of such content loses its genuineness. Therefore, to tackle the problem of deepfakes when it comes to a case before any court of law, precedents shall play a very crucial role.

The technology of deepfakes is a critical problem to tackle, involving the application of law, technology and ethics requires an evolving approach. The fact that technology is continuously evolving, it is important for India to keep up with these developments and strengthen its legal foundation to address issues pertaining to technology. Deepfakes are one of a kind but as time goes on, new AI-generated content might become available. Therefore, it is crucial to reinforce our legislature and enact strict regulations to address the issue of deepfakes, to ensure that our nation does not fall behind in the future.