VOLUME 4 | ISSUE 3

INTERNATIONAL JOURNAL OF ADVANCED LEGAL RESEARCH

DEEPFAKE TECHNOLOGY AND HUMAN RIGHTS: UNMASKING THE ETHICAL CHALLENGES AND IMPACTS

- Jayalakshmi V[1]

Abstract:

Deepfake technology, fuelled by advancements in artificial intelligence and machine learning, has emerged as a formidable tool for manipulating audiovisual content, raising concerns about its potential impact on various facets of society.Deepfake technology has not only transformed the landscape of digital content but has also raised critical concerns regarding its implications for human rights. Deepfake technology presents a multifaceted challenge to human rights across various dimensions. Privacy violations are rampant as individuals face the unauthorized use of their facial images and voices, infringing upon the right to control one's personal information. The technology's capacity to generate highly realistic yet entirely fabricated content exacerbates concerns about freedom of expression, enabling the spread of misinformation that can manipulate public opinion and potentially silence dissenting voices. Furthermore, deepfake manipulation can perpetuate discrimination and targeted harassment, amplifying harmful stereotypes and harming the reputation of individuals. The psychological impact on victims cannot be understated, as being subject to falsified digital content can lead to reputational damage, strained relationships, and profound emotional distress. Effectively addressing the human rights implications of deepfake technology necessitates a comprehensive and adaptive approach, encompassing technological innovations, legal adaptations, ethical considerations, and widespread public awareness efforts. Only through such a holistic strategy can societies hope to mitigate the multifaceted negative consequences posed by deepfake advancements.

Keywords: Deepfake, audiovisual, digital content, human rights, harassment, reputation

[1] Assistant Professor; Mother Teresa Law College, Pudukottai.

©2024 International Journal of Advanced Legal Research

Understanding the concept of deep fake Technology:

According to the Merriam-Webster Dictionary, deep fake means "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."[2]

Deepfakes are generated through a process known as Generative Adversarial Networks (GAN), wherein two neural networks compete against each other. One of these networks, the Generator, is responsible for creating images or videos to make them appear as accurate as possible. The different network, the Discriminator, assesses the content to determine whether it is authentic or produced by the Generator.

This technology gained prominence in the mid-2010s as advancements in deep learning techniques and increased computational power made it more accessible. Initially, deepfakes were mainly associated with creating realistic face swaps in adult content, but their application has expanded to include political manipulation, misinformation, and entertainment.

A deepfake refers to a specific kind of synthetic media where a person in an image or video is swapped with another person's likeness.This technology learns by analyzing a large amount of data, often pictures or videos of a particular person, and then uses that knowledge to generate new content that appears authentic. For instance, it can make videos of people saying or doing things they never actually did.

In 2020, the Bharatiya Janata Party employed manipulated videos featuring its president criticizing the opposing party in its campaign for the Delhi election. This marked the initial instance of a political party using deepfake technology in India.[3] However, Deepfake technology has entertaining applications, like special effects in movies; there are concerns about its potential misuse, such as creating deceptive content for malicious purposes.

[2]Webster N, 'Deepfake'<https://www.merriam-webster.com/dictionary/deepfake> accessed 23 November 2023

[3]Binayak Dasgupta, 'BJP's deepfake videos trigger new worry over AI use in political campaigns' (*Hindustan Times* New Delhi, 20 February 2020)<https://www.hindustantimes.com/india-news/bjp-s-deepfake-videos-trigger-new-worry-over-ai-use-in-political-campaigns/story-6WPlFtMAOaepkwdybm8b1O.html>accessed 23 November 2023

©2024 International Journal of Advanced Legal Research

Benefits of Deepfake Technology

Positive uses of deepfake technology benefit society as long as it is lawful and ethical. It has enormous beneficial potential in commercial applications.Employing deepfake technology can enhance a teacher's ability to conduct captivating lessons, surpassing the confines of conventional visual and media formats.The utilization of artificial intelligence-generated synthetic press has the potential to animate historical figures, providing an immersive educational experience within the classroom.

Additionally, it can emerge as a potent resource for independent storytellers, offering significant capabilities at a fraction of the usual expenses. Deepfakes are a valuable tool for authentically embodying the fundamental elements of comedy or parody, encompassing reflection, exaggeration, distortion, and the creative reinterpretation of actual events.[4]

Deepfake technology extends its utility by concealing individuals' voices and faces, safeguarding their privacy.People can employ Deepfakes to craft avatar experiences for self-expression online, empowering them to independently convey their values, ideas, and beliefs through a personalized digital representation. This is particularly beneficial for individuals, including those with physical or mental disabilities, seeking to express themselves effectively online. Deepfakes provide individuals with innovative tools for self-expression and seamless integration into the digital landscape, fostering new avenues for personal empowerment in the online sphere.

Deepfake Technology issues:

The primary danger associated with the inappropriate use of deepfake videos is the potential for creators to manipulate a video featuring someone else, using this technology to insert false statements and mislead the audience.The risk extends to privacy invasion through the unauthorized creation of explicit or compromising content featuring individuals.

Deepfakes can also be leveraged in social engineering attacks and identity theft, eroding trust in digital media and making it challenging for people to discern between authentic and manipulated content. Addressing these privacy concerns requires the

[4] 'Applications of Deepfake Technology: Positives and Dangers' (*knowledge Nile*) < https://www.knowledgenile.com/blogs/applications-of-deepfake-technology-positives-and-dangers> accessed 30 November 2023

©2024 International Journal of Advanced Legal Research

development of effective detection methods, heightened awareness, education about the risks, and potential adjustments to legal frameworks to safeguard individuals from the misuse of deepfake technology.

Deepfakes can create realistic-looking videos or audio recordings that convey false information or messages. This can contribute to the spread of misinformation and disinformation, potentially leading to public confusion and mistrust.

Deepfake technology can be employed in financial scams, where criminals use fake audio or video recordings to impersonate executives or other trusted individuals within organizations. This could lead to fraudulent transactions or disclosure of sensitive information.

Deepfakes can be used as tools in cyber warfare and espionage, posing a threat to national security. Fake videos or audio recordings could be created to sow discord, manipulate public sentiment, or compromise the reputation of political leaders.

These examples illustrate the diverse applications of deepfake technology, ranging from entertainment and satire to the potential misuse of disinformation.

- A deepfake video featuring Facebook CEO Mark Zuckerberg emerged in 2019, where he appeared to be discussing Facebook's control of people's data. The video was created to draw attention to the issue of misinformation and deepfake technology.[5]
- A deepfake video featuring former President Barack Obama was created by BuzzFeed and Jordan Peele to raise awareness about the dangers of deepfake technology. The footage showed Obama delivering a fabricated public service announcement.[6]
- In 2021, a deepfake artist, Chris Ume, created highly realistic videos of Tom Cruise impersonations and posted them on TikTok. The videos raised concerns about the potential misuse of deepfake technology for impersonation.[7]

[5]Jordan Smith, 'Artist tests Facebook video policy with 'deepfake' of Mark Zuckerberg' (*CNBC*, 12 June2019)<https://www.cnbc.com/video/2019/06/12/an-artist-just-made-a-deepfake-of-mark-zuckerberg.html> accessed 27 November 2023

[6]James Vincent, 'Watch Jordan Peele use AI to make Barack Obama deliver a PSA about fake news' (*The verge*, 17 April 2018) <https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed> accessed 27 November 2023

[7]Marcia Sekhose, 'The guy who created Tom Cruise deepfake videos on TikTok now owns an AI company' (*Business Insider INDIA*, 09 August 2021) <https://www.businessinsider.in/tech/news/chris-um-who-created-tom-cruise-deepfake-videos-on-tiktok-now-owns-an-ai-

©2024 International Journal of Advanced Legal Research

- In 2018, China's state-run news agency, Xinhua, unveiled an AI-powered virtual news anchor. While not a traditional deep fake, it raised concerns about using AI to simulate realistic-looking news presenters.[8]

Deepfake detection strategies and technologies:

There is a constant race between deepfake creators and detection developers to adapt to evolving methods.Detecting deepfake technology is like using a digital detective to spot fake videos or pictures. One way is by paying attention to small details that might look slightly off, like strange movements or unnatural facial expressions.

Another trick is to listen closely to the voices in videos—if something sounds a bit weird or not quite right, it might be a clue that it's a fake. Sometimes, technology can help by using intelligent computer programs that have learned from lots of real videos and pictures to tell when something doesn't seem genuine. So, it's like having a bunch of digital detectives working together to catch the fakes.

As new deepfake techniques emerge, detection systems may not be equipped to identify "zero-day" threats immediately.Zero-day threat is dangerousbecause its existence is only known to the attacker.

In identifying deepfakes, a significant indicator is the inadequate synchronization of facial expressions when faces are exchanged, constituting a prominent high-level semantic feature in detection methods. A multimodal strategy for detection involves identifying audio-visual disparities, making it applicable to deepfakes encompassing both facial and audio alterations. Another multimodal technique involves leveraging spatial-temporal features, including scrutinizing visual anomalies within specific video frames (intra-frame inspection) and assessing temporal patterns across various video sequences (inter-frame examination).[9]

The limited availability of diverse training data hampers the effectiveness of detection models across various deepfake variations. Additionally, the natural variability in human

company/articleshow/85170929.cms> accessed 27 November 2023

[8]Lily Kuo, 'World's first AI news anchor unveiled in China' (*The Guardian*,09 November 2018) <https://www.theguardian.com/world/2018/nov/09/worlds-first-ai-news-anchor-unveiled-in-china> accessed 27 November 2023

[9] Naitali and others, 'Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions',*MDPI*2023, 12(10) <https://doi.org/10.3390/computers12100216>accessed 12 December2023

©2024 International Journal of Advanced Legal Research

expression and the potential for ethical concerns, such as privacy issues and false accusations, add complexity to the detection process. The use of adversarial attacks on detection algorithms by deepfake creators further complicates the development of foolproof detection methods. Striking a balance between effective detection and ethical considerations and creating user-friendly tools for non-experts remains challenging.

Human Rights Violations in Deepfake Technology:

The deepfake technology was initially developed for applications in entertainment and education, as it can generate AI-generated characters for specific purposes without the need to employ human beings for the same. However, its usage has gradually taken a different turn, and its purpose has changed.Individuals can become victims of privacy invasion by creating deepfake content that depicts them engaging in activities they never did. Deepfake technology has been used to make explicit content by superimposing an individual's face onto the bodies of actors in adult content. This type of non-consensual creation and distribution of explicit material can be a severe violation of privacy and may lead to emotional distress and reputational harm.

Deepfakes can be used for identity theft by creating fabricated videos or audio clips impersonating individuals. This can lead to the misuse of personal information, financial fraud, or other malicious activities compromising an individual's privacy.

Deepfakes raise concerns about unauthorized surveillance and stalking. The ability to create realistic videos depicting someone in various locations or situations without their knowledge or consent can violate their privacy rights and develop a sense of constant surveillance.

Deepfake pornography represents a disturbing assault on personal privacy, particularly affecting women. This reprehensible practice constitutes a blatant violation of individuals' rights, with a pronounced impact on women who often bear the brunt of such malicious attacks. Deepfake pornography is not only a violation of privacy but also a cruel exploitation of individuals, preying on their most intimate moments to satisfy malicious intentions. Mostly, Actors are affected by this deepfake pornography.

The process often begins with collecting a substantial amount of facial or body imagery of the target individual. This can be obtained from publicly available sources, such

©2024 International Journal of Advanced Legal Research

as social media profiles or private sources, if the attacker has unauthorized access to personal data.Once the deepfake model is trained and capable of accurately replicating the target person's likeness, it can be used to generate new content. This content often appears as if the target individual is actively participating in explicit or pornographic scenarios.In India, Bollywood actresses such as Priyanka Chopra, Kajol Devgan, and Rashmika Mandana have unfortunately been targeted as victims of deepfake pornography.[10]

Revenge porn can be sighted in India in 2018. A famous Indian journalist, Rana Ayyup, was threatened, harassed, and humiliated by swapping her face onto a pornographic clip and making it appear as if she acted in that clip, though she had not. [11]

Legal frameworks of deep fake technology:

Like many other technologies, deepfake technology also has two sides. However, the challenge here is the thin line in distinguishing the original from the fake. In most cases, the fake is often superior to the original. Hence,strict legal frameworks should be needed to combat deepfake offenses as their effects endanger and threaten human lives and privacy.

In the US, a separate Act has been enacted for this purpose. It introduced the Bill in 2019 and became an Act in 2023. The Act is the "Defending Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2023" or the "Deepfakes Accountability Act."[12] This Act requires the disclosure of the identity of the deepfake creators.[13] Whoever fails to comply with the requirements under the Act is subjected to a fine and imprisonment of up to 5 years. It also establishes civil penalties and permits individuals to bring civil actions for damages.[14]

[10]Bidisha Saha,'Let alone Rashmika Mandana, Internet is filled with deepfake Bollywood porn'(*India Today* 07 November 2023) <https://www.indiatoday.in/india/story/let-alone-rashmika-mandanna-internet-is-filled-with-deepfake-bollywood-porn-2459404-2023-11-07> accessed 27 November 2023.
[11]Nisha Dhanraj Diwani and others,*Handbook of Research on Cyber Law, Data Protection, and Privacy*(1st edn, IGI Global Publisher of Timely Knowledge,2022,)
[12] H.R.5586 - DEEPFAKES Accountability Act<https://www.congress.gov/bill/118th-congress/house-bill/5586/text?s=2&r=1&q=%7B%22search%22%3A%22deepfakes%22%7D> accessed December 6, 2023
[13](TRANSPARENCY REQUIRMENTS) of the DEEPFAKES Accountability Act 2023, s 2
[14] (Advanced technological false personation record) of the DEEPFAKES Accountability Act 2023, s 1041

©2024 International Journal of Advanced Legal Research

China's new deepfakes law, "Provisions on the Administration of Deep Synthesis of Internet Information Services," took effect on January 10, 2023.[15] According to China's regulations relating to deepfakes, usersmust provide approval before their image is utilized in any deep synthesis technology. The use of the technology by deep synthesis services for spreading false information is not permitted. Deepfake services need to verifyusers' genuine identity. Any synthetic content should include a notification indicating that the image or video has been altered using technology. Prohibited content includes content that violates existing laws, poses a threat to national security and interests, harms the national image, or disrupts the economy.[16]

The legal framework in England about AI and the rights of individuals remains largely unsettled. The current laws find it challenging to keep up with the swiftly evolving technological landscape, exposing individuals to potential exploitation. Although the impending Online Harms Bill in England and Wales aims to establish a criminal offense for sharing "deepfake" pornography, it does not extend to prohibiting other forms of AI-generated content created without the subject's consent. Consequently, individuals are reliant on existing laws in such situations.[17]

So far,India does not have specific legislation addressing deepfake offenses. However, certain existing laws can be invoked to address related concerns. A few provisions of The Information Technology Act of 2000 address deepfake cybercrimes. The recently enacted Digital Personal Data Protection Act 2023 also has dealt with the misuse of personal data.

Section 66E of the IT Act 2000[18]explicitly addresses the unauthorized capturing or publishing of private images of an individual without their consent. It considers such acts as offenses and provides for legal consequences.This provision aims to protect individuals from the unauthorized and intrusive capture, publication, or transmission of private images, emphasizing the importance of privacy in the digital realm. Hence, deepfake technology,

[15]Afiq Fitri, 'China has just implemented one of the world's strictest laws on deepfakes' (*TECH MONITOR* January 10, 2023) <https://techmonitor.ai/technology/emerging-technology/china-is-about-to-pass-the-worlds-most-comprehensive-law-on-deepfakes> accessed on December 6, 2023

[16]Arjun Kharpal,'China is about to get tougher on deepfakes in an unprecedented way. Here's what the rules mean'(*CNBC*, December 22, 222)<https://www.cnbc.com/2022/12/23/china-is-bringing-in-first-of-its-kind-regulation-on-deepfakes.html> accessed on December 6, 2023

[17]Oliver Lock, 'The legal issues surrounding deepfakes and AI content'(*FARRER & Co,* October 12, 2023) <https://www.farrer.co.uk/news-and-insights/the-legal-issues-surrounding-deepfakes-and-ai-content/> accessed on December 7, 2023

[18] The Information Technology Act 2000, s 66E

©2024 International Journal of Advanced Legal Research

which manipulates and misrepresents individuals,falls under this section's purview. Violations of this section result in imprisonment of up to 3 years or a fine of up to 2 lakhs or both.

Section 66D of the IT Act 2000[19] addresses instances where individuals engage in fraudulent activities by assuming another person's identity through electronic means.This provision is designed to curb cybercrimes related to identity theft and fraudulent impersonation through electronic devices or computer resources. Offenders who engage in cheating by adopting the identity of another person using digital means may face imprisonmentfor up to 3 years and a fine of up to 1 lakh under this section.Additionally, the Copyright Act of 1957provides copyright protection against unauthorized use of works, allowing copyright owners to take legal action.[20]

According to the DPDP Act 2023, the data fiduciary shall use the personal data of an individual who is otherwise called a Data Principal only with their consent and only for legitimate purposes.[21]

Addressing legal hurdles posed by deep fake technology:

Since offenses occur in cyberspace, tracking the source of the crime poses a significant challenge. The generation of fraudulent content might occur outside the jurisdiction of the victim, spanning national or regional borders. Due to disparities in legal frameworks among jurisdictions, perpetrators may evade appropriate punishment, highlighting the need for a more cohesive international approach to address these challenges.

The growing use of deep fakes as fabricated evidence in legal proceedings raises serious concerns for the rule of law. This trend may result in prolonged trials as evidence authenticity is disputed, creating uncertainties. The risk of courts inadvertently accepting manipulated evidence further complicates the judicial process. Convicted individuals might publicly proclaim innocence based on the alleged use of fake evidence, and even if innocence is later established, the potential harm to one's reputation and life is significant. Moreover, in cases involving sensational news and debunked fake stories, the initial misinformation often

[19] The Information Technology Act 2000, s 66D
[20]The Copy Right Act 1957, s 51
[21]The Digital Personal Data Protection Act 2023, s4

©2024 International Journal of Advanced Legal Research

garners more attention than the subsequent correction, leaving a lasting impact on public perception.[22]

The dynamic progression of deepfake technology exacerbates difficulties for automated detection systems, resulting in heightened challenges, especially in the presence of contextual intricacies. This represents a notable menace to legal processes, potentially extending trial durations and amplifying the potential for erroneous assumptions.[23]

Suggestions:

Reducing deepfake issues requires a multifaceted approach involving technology, policy, education, and collaboration. Here are some suggestions:

- Research and development of advanced detection tools and algorithmsmust be needed to identify and flag deepfake content across various platforms. Collaboration between tech companies, research institutions, and government agencies can accelerate progress.
- Educate the public about the existence of deepfakes, such as how they are created and their potential impact. Encourage critical thinking skills and teach individuals how to identify and evaluate the authenticity of online content.
- Develop and enforce laws and regulations that address the creation, distribution, and misuse of deepfake technology. This includes laws related to intellectual property, defamation, privacy, cybersecurity, and election integrity.
- Promote transparency in using deepfake technology by requiring clear labeling or disclosures when content has been digitally manipulated. Hold individuals and organizations accountable for the creation and dissemination of malicious deepfakes.
- Foster international cooperation and coordination to address the global nature of the deepfake challenge.

Conclusion:

[22] Bart van der Sloot & Yvette Wagensveld, 'Deepfakes: regulatory challenges for the synthetic society', *Science Direct*(2022) 46(105716) <https://doi.org/10.1016/j.clsr.2022.105716> accessed 13 December 2023

[23]Harshavardhan Mudgal, 'The deepfake dilemma: Detection and decree' (*Bar and Bench*, November 18, 2023) < https://www.barandbench.com/columns/deepfake-dilemma-detection-and-desirability> accessed 11December2023

©2024 International Journal of Advanced Legal Research

In conclusion, while technology has historically aimed to benefit humanity, its ethical application remains pivotal in determining its impact. Deepfake technology, like many innovations, holds potential for advancement and efficiency, yet its misuse threatens fundamental human rights, notably privacy. As the digital landscape evolves, legislative frameworks must adapt to address emerging challenges. Although strides have been made, such as India's enactment of the DPDP Act 2023, further stringent measures are imperative to safeguard against the pernicious effects of deepfakes. In December 2023, the Indian government issued an advisory to all social media and internet intermediaries, including WhatsApp, Instagram, Facebook, and Google, seeking strict compliance with the existing IT rules, specifically targeting growing concerns around misinformation powered by deepfake technology.[24]This underscores the urgency of this matter, signaling a collective effort to combat misinformation and uphold online integrity. As we navigate this ever-changing technological terrain, proactive measures and international collaboration are essential to mitigate the risks posed by deepfake technology and ensure a secure digital future for all.

[24]'Deepfake menace: Govt issues advisory to social media platforms to comply with IT rules' (*The Economic Times*, 27 December 2023) <https://economictimes.indiatimes.com/tech/technology/deepfake-menace-govt-issues-advisory-to-intermediaries-to-comply-with-existing-it-rules/articleshow/106297813.cms> accessed on 02 February 2024

©2024 International Journal of Advanced Legal Research