Page 2267 - 2274 **DOI:** *https://doij.org/10.10000/IJLMH.111463*

INTERNATIONAL JOURNAL OF LAW MANAGEMENT & HUMANITIES

[ISSN 2581-5369]

Volume 4 | Issue 4

2021

Deconstructing Deepfake: Tracking Legal Implications and Challenges

MADHURA THOMBRE[1]

ABSTRACT

'Science gathers knowledge faster than society develops wisdom!'

-Isaac Aminov

Technological developments have touched every corner of our life and has provided us with countless opportunities. Changing technology is influencing the structure of legal ecosystem and being a double-edged sword; it comes with several potential threats. Advances in the field of artificial intelligence have made generation of fake videos, images quite easy. Deepfake use a form of artificial intelligence known as deep learning to allow users to manipulate images, videos, audios and even allowing them to create footage of events which never happened in reality. The technology has several benefits in the fields of entertainment, filmmaking, media, marketing, health and so on. However, it also comes with certain risks. Malicious use of deepfake could stifle its valuable utilities. It has a potential to completely alter the cybersecurity landscape by bringing in the risks ranging from privacy violation, abuse, defamation to breach of public peace, election manipulation ultimately threatening the national & international safety and order. The article is an attempt to take an overview of deepfake technology along with its legal implications and associated risks.

I. INTRODUCTION

'Forget artificial intelligence - in the brave new world of big data, it's artificial idiocy we should be looking out for."

—Tom Chatfield

Synthetic media or Artificial Intelligence generated media is a rapidly growing field with deepfake emerging as its prominent form. It is allowing individuals to do or say things which they have never done or said thereby creating a convincing but completely fabricated content. Deepfake is a powerful tool which can be used for creative effects as well as for exploitation & disinformation. Being a double edged sword, it has several legal implications in social

[1] Author is an Assistant Professor at School of Business and Law Navrachana University, Vadodara, Gujarat, India.

 [ISSN 2581-5369]

domains.

II. ORIGIN AND MEANING OF DEEP FAKE

Deepfake is an amalgamation of two terms-'deep learning' and 'fake'[2]. Deepfake refers to a specific form of synthetic media where artificial intelligence is used for the purpose of altering an image or an audio-video content to make it look real but actually it has never happened. By manipulating images, videos, and voices of real people, deepfake can portray someone doing things they never did, or say things they never said[3]. The video/audio/photo that it creates is too realistic to detect the manipulation. At the base of it is a 'Deep learning neural network'. Deep learning is a subset of machine learning capable of handling multi-layer neural networks. A person whose image or facial expression is to be manipulated; thousands of his images are feed to the software. Machine learns the images, analyses it and accordingly prepares an algorithm as to how a person's facial expressions would look like for a particular content in order to make mimicking feel real. Similarly, the voice, accent, tone is also learnt and is used to perfectly replace the original one.

Generative adversarial network(GAN) is a way to further improvise deepfake and make it more believable[4]. The algorithms generated by artificial intelligence goes through countless number of feedbacks. After multiple rounds of detection of patterns and subsequently removing flaws; deepfake content is improvised. Consequently, a fabricated but incredibly realistic image/video/voice is made to replace the original content or make one even of a non-existing person or event.

Deepfake can be said to have its origin in December 2017 where a Reditt user coined the term for videos of celebrities engaged in pornographic activities[5]. The videos were looking quite realistic, but were actually fake ones created by using the 'deep learning' technology. Artificial intelligence tool used images, videos of celebrities to generate an algorithm to study the patterns of facial expression, voice and with multiple layers of detection and improvement, it created a nearly perfect image, voice and expression of celebrities to synthesize it with the original actors in pornographic content. The videos went viral on pornography websites and created a lot hue & cry by bringing public humiliation to such celebrities.

[2] *Deepfakes are the evolution of fake news and are equally dangerous*, STE DAVIES (July 7, 2021, 03.00 PM) https://www.illinoislawreview.org/blog/ai-deepfakes/

[3]Meredith Somers, *Deepfakes, explained*, MITSLOAN (July 7, 2021, 04.00 PM) https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained

[4] Grace Shao, *What 'Deepfakes' are and how they may be dangerous*, CNBC (July 7, 2021, 03.30 PM) https://www.cnbc.com/2019/10/14/what-is-deepfake-and-how-it-might-be-dangerous.html

[5]Ian Sample, *What are deepfakes and how can you spot them*, THEGUARDIAN (July 7, 2021, 04.15 PM) https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them

 [ISSN 2581-5369]

Deepfake is not limited only to pornography. Lot of non-pornographic content, where faces of politicians, comedians swapped with the original faces, went viral on the internet. These videos showed people doing or saying things which they never did. Barack Obama's deepfake video where he is calling the then US President Donald Trump 'complete dipshit' or deepfake video of Mark Zuckerberg claiming 'complete control over people's data' went viral on the internet and before majority of the people could realize that it was deepfake, it already caused a damage that it intended to.

Deepfake voice recordings also caused havoc. Belgian Prime Minister linked corona virus pandemic to climate change in a recorded speech which was later found out to be a manipulated one[6]. In March 2019, CEO of a UK subsidiary of a German Energy firm got a call from his German chief for transfer of 2 lakh pounds into a Hungarian bank account. Later it was discovered that a fraudster used artificial intelligence to mimic the voice, German accent and the tone of a German chief to create a deepfake recording for siphoning out money from the UK based firm. This new technology of generating nearly realistic content is making it alarmingly difficult to distinguish between genuine and fake content and can have potential socio-legal implications.

III. IMPLICATIONS AND LEGAL ISSUES

Implications of deepfake can be multifold and it depends upon the way in which it is put to use. The term deep fake has negative connotations, but it can be used for beneficial purposes in marketing, entertainment field. Online gaming industry is experimenting deepfake to use 'artificial intelligence generated voice skins' to enhance experience of the gamers. With face cloning, voice cloning, the entertainment & advertising industry can utilize the potential of this technology to its best. Several stakeholders have started using the term 'artificial intelligence –generated synthetic media' instead of deep fake in order to eclipse the negative connotation attached to deepfake.

Disney has revealed that it is using deepfake to enhance the quality of its audio-visual content and such experiments are already undertaken in the filmmaking industry. Filmmakers have digitally inserted the archival footage of John Kennedy and manipulated his mouth movements. Synthetic voice media and image media combined with artificial intelligence was used to digitally insert late actor Paul Walker face and voice in 'Furious 7' during which the actor had died[7]. During the worldwide Covid-19 pandemic, when shooting became difficult, some

[6] Meredith Somers, *Deepfakes, explained,* MITSLOAN (July 7, 2021, 04.00 PM) https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained

[7] *Deepfake videos: Inside Pentagon's race against deepfake*, CNN BUSINESS (July 7, 2021, 02.00 PM),

 [ISSN 2581-5369]

companies did start looking for deepfake technology for creation of training videos. Deepfake technology is also being tested for detection of tumours in healthcare field. Hence it can be said that technological advancements in deepfake can be used for beneficial purposes and ultimately it depends upon how it is used.

Technology however, is a double-edged sword and the risks associated with it cannot be ignored. Deepfake is much more than a mere face-swapping or Photoshop or a voice replacement and it has several serious socio-legal implications. With increasing internet penetration and reducing digital divide across the world, the accessibility of digital content has enhanced fourfold and consequently the set of associated risks as well. The widespread accessibility of technology makes it available practically to anyone-individual, state as well as non-state actors which may lead to catastrophic consequences.

IV. PRIVACY VIOLATION, ABUSE & FRAUDS IN CYBERSPACE

Technology of Deepfake allows one to manipulate the video thereby swapping the original face or voice with anyone else's face or voice or image. Images of politicians, celebrities or even of any common man that are available across the internet, social media networks can be easily accessed by the perpetrators without consent of a concerned person. The audio-visual content so generated can bring irreparable harm to a person's reputation and it can have potential implication on one's personal and professional life. Pornographic videos of celebrities which were later discovered as 'Deepfake' had already been widely circulated across the globe and caused damage of unimaginable magnitude to the reputation of these celebrities. 'Revenge porn' is a most common form of abuse via deepfake technology. It blatantly violates one's basic human right to privacy and right to reputation. Malicious use of deepfake is changing the traditional landscape of cyber security and is being deployed for bringing public humiliation to the victims and for frauds.

A research by regmedia has found out that deepfake technology has significant online presence with 8 out of top 10 pornographic websites have reportedly hosting deepfakes. Significant number of women, celebrities is already affected by deepfake pornography. Deepnude, a website and computer app in which a computer algorithm was trained to create deepfake synthetic images of women by allowing users to virtually remove clothes and generate naked body. After receiving stark criticism; the website was taken down but by then, it had 5,45,000 visits within 24 hours and even after taking down, the software has continued to be

https://edition.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/

 [ISSN 2581-5369]

independently repackaged & distributed through open source repositories and dark web[8].

Abusive deployment of deep fake can open Pandora box of troubles in every field. Attempts of fraud, deceit, espionage through deepfake videos, profiles, and images are on rise with increasing accessibility of the technology. In March 2019, a social media profile of a journalist employed with Bloomberg- 'Maisey Kinsley' appeared and deceived many short sellers of Tesla by demanding their personal information[9]. There was no such person employed with Bloomberg in reality. It was found out to be a synthetically generated imagery and profile. It was subsequently blocked by Twitter and LinkedIn. A LinkedIn account of 'Katie Jones' posing as researcher from a US think tank, was discovered to have been generated by using deepfake. But before that it had contacted several government officials in between March-April 2019 for secret information[10]. It was only when some visual anomalies of the profile and photos were discovered by a few, experts investigated the matter and raised an alarm as an attempt of espionage from some foreign country. A UK based subsidiary of a German company was defrauded by a deepfake generated voice impersonation of a CEO to demand urgent money transfer of around 2 lac pounds to a supplier. Commodification of deepfake is giving rise to novel threats in the cyberspace landscape[11].

The producers and distributors of deep fake will certainly invite legal implications including that of the charges of defamation and sometimes of copyright violation. Information technology related laws or data protection laws in many countries including India, Scotland, USA have a specific criminal offence for sharing intimate or private images without consent via internet. However, the online presence of deepfake technology is rapidly expanding encompassing the borders & making it a global concern. In case of trans-natural nature of these crimes, jurisdiction issue is a major challenge that demands a streamlined approach across the globe. Available legal options of copyright protection, privacy violation under cyberspace regulations or actions defamation have limited utility and they may not serve the purpose fully. Once the malicious deepfake video is on the internet, it is extremely difficult to wipe out all the copies of the same from digital space.

---

[8] Henry Ajder, Giorgia, Patrini, Franscesco Cavalli, Laurence Cullen, *The State of Deepfake: Landscape, Threats and Impact*, REGMEDIA (July 7, 2021, 04.30 PM) https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

[9] Glenn Fleishman, *How to spot the realistic fake people creeping into your timeline*, FASTCOMPANY (July 7, 2021, 05.00 PM) https://www.fastcompany.com/90332538/how-to-spot-the-creepy-fake-faces-who-may-be-lurking-in-your-timelines-deepfaces

[10] Raphael Satter, *Experts: Spy used AI-generated face to connect with targets*, APNEWS (July 7, 2021, 02.15 PM) https://apnews.com/article/ap-top-news-artificial-intelligence-social-platforms-think-tanks-politics-bc2f19097a4c4fffaa00de6770b8a60d

[11] Meredith Somers, *Deepfakes, explained*, MITSLOAN (July 7, 2021, 04.00 PM) ,https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained

 [ISSN 2581-5369]

V. MISINFORMATION

Images, audio-video recordings have been a vital source of information and are considered as evidences for the events. Techniques of manipulation of the content have been in use for decades; however, the technology of deepfake could be a complete game-changer. As it develops and proliferates, anyone could've the ability to make a convincing fake video, including some people who might seek to 'weaponize' it for malicious purposes[12].

Facial mapping, audio synchronising techniques coupled with artificial intelligence has made creation of fake content very simple and that can destabilise the social harmony, disrupt societal trust and cause havoc by spreading misinformation. Fake contents is seen to have been used for political sabotage, blackmailing, bullying and now with more sophisticated techniques in hand; the threat of malicious falsehood has increased fourfold. Deepfake videos of Barak Obama ridiculing the then President Donald Trump or that of Mark Zuckerberg claiming 'complete control over people's data through Facebook' have caused havoc, panic amongst people due to hyper-realistic content. Making deepfake videos of politicians and public figures is way easier as lot of their photos, videos are easily available on the internet. A few such deepfake videos of politicians getting circulated through social media can spread plethora of misinformation resulting into severe consequences ranging from confusions, uncertainty, eroding public trust, undermine democratic set-up through election manipulation.

Deepfake is no longer a part of fiction stories and the risks are no more hypothetical. Gabon, a small central African oil-rich country has witnessed the implications and chaotic scenario due to a mere suspicion of use of deepfake in a video message of the President. President Ali Bongo Ondimba was reportedly hospitalized in 2018 after attending a summit in Saudi Arabia and was not heard for a couple for months since then. Rumours of his death and severe illness started spreading very fast and meanwhile he appeared in a video delivering new year address towards the end of 2018. Rumours that the video is fabricated and is made with deepfake technology engulfed the whole country and in the event of this uncertainty, Gabonese soldiers attempted coup in January 2019[13]. It took several months to bring back the social stability and order. Ruling Bhartiya Janata party of India used Deepfake technology to alter original campaign footage in English of candidate Manoj Tiwari to alter his intonation, words and

[12]*Deepfake videos: Inside Pentagon's race against deepfake*, CNN BUSINESS (July 7, 2021, 02.00 PM), https://edition.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/

[13] Sarah Cahlan, *How misinformation helped spark an attempted coup in Gabon*, WASHINGTONPOST (July 8, 2021, 11.00 AM) https://www.washingtonpost.com/politics/2020/02/13/how-sick-president-suspect-video-helped-sparked-an-attempted-coup-gabon/

 [ISSN 2581-5369]

speech into Hindi language to make it more appealable to voters[14]. Earlier in December 2018, researchers in UK made a deepfake video of two candidates endorsing each other. But it was merely for the purpose of awareness creation about existence of deepfake and not for garnering votes. Indian political party became the first ever to use deepfake technology for campaigning purposes[15].

The breath-taking pace with which the technology is developing, it can now be used for building convenient image of politicians, discredit the opponents, cause irreparable damage to someone's reputation and ultimately influencing the politics like never before. It is a situation which can be exploited by every state and non-state actor for malicious purposes and can trigger a new era of arms race amongst them.

VI. DEVALUATION OF TRUTH

Deepfake can pose a serious threat to authenticity of audios, videos presented as evidences everywhere including the courts. Mere suspicion of use of deepfake can create confusion, chaos and can cause serious devaluation of truth. In a world where deepfake content is becoming common, it is easier for anyone to discredit the genuine content as fake. Moreover, it will lead to a phenomenon known as 'the liar's dividend' where every single video would be looked upon with suspicion and will create a big trust deficit between public institutions and common people. Research by Brooking institutions has pointed out that the hanging uncertainty about authenticity of every audio, video or images could lead to catastrophic consequences and can completely end our trust on institutions[16]. The widespread accessibility of technology makes it vulnerable to be exploited by any state or non-state agents against the rival and it can undermine public safety or even national security at large. A sex tape featuring Cabinet Minister of Malaysia and his rival politician's aide engaging in same-sex activity went viral[17]. It was quickly dismissed as being made with deepfake technology. However, it did cause damage of creating suspicion, destroyed the credibility and was seen as an attempt to crumble the ruling coalition leading to further worsening of the rivalries[18]. Belgium witnessed the

[14] Charlotte Jee, *An Indian Politician is using deepfake technology to win new voters*, TECHNOLOGYREVIEW (July 8, 11.15 AM) https://www.technologyreview.com/2020/02/19/868173/an-indian-politician-is-using-deepfakes-to-try-and-win-voters/

[15] ibid

[16] William Galston, *Is seeing still believing? The deepfake challenge to the truth in politics*, BROOKINGS (July 8, 11.30 AM, 2021) https://www.brookings.edu/research/is-seeing-still-believing-the-deepfake-challenge-to-truth-in-politics/#cancel

[17] Nic Ker, *Is the political aide viral sex video confession real or deepfake?*, MALAYMAIL (July 8, 2021, 11.45 PM) https://www.malaymail.com/news/malaysia/2019/06/12/is-the-political-aide-viral-sex-video-confession-real-or-a-deepfake/1761422

[18] Ibid

atmosphere of overall public confusion when a deepfake video of Belgian Prime Minister where she claimed that the recent pandemics of SARS, Ebola and Covid-19 are directly linked to the destruction caused by humans to the natural environment went viral[19]. It was tough for experts to conclusively prove the use of deepfake technology in a convincingly realistic video.

In a world where seeing is no longer believing, the ability of a large community to agree on what is true-much less to engage in constructive dialogue about the same-seems precarious[20].

VII. CONCLUSION

Deepfake technology has a potential of being a game-changer for several fields such as media, entertainment, marketing and many more. However, malicious use of deepfake can have disastrous consequences of unimaginable magnitude with a real risk to national & international order. A new arms race may trigger to further exploit the technology or even weaponize it to achieve malicious & vested interests. Technological advancements in deepfake are outplaying our capability and efforts to detect them. Complete ban on deepfake generated content can have chilling effect on free speech. Moderating its usage and protecting humanity from the associated risks through legal solutions should be the focus of our efforts to combat dangers of deepfake. Transnational nature of cyberspace and lack of harmonised laws on data protection are two major hindrances to the path of deepfake regulation. Researchers are working on development of Artificial intelligence based deepfake detection techniques and a handful of start-ups are offering software to detect deepfake. However, mere technological remedies are not enough. We must undertake a critical review of existing legal remedies in light of the legal and ethical implications of deepfake. Legislative efforts to protect from deepfake risks coupled with campaigns for awareness, global streamlined approach to combat deepfake dangers is need of an hour.

*****

[19] *Belgium posts deepfake of Belgian premier linking covid-19 with climate crisis*, BRUSSELSTIMES (July 8, 2021, 12.15 PM) https://www.brusselstimes.com/news/belgium-all-news/politics/106320/xr-belgium-posts-deepfake-of-belgian-premier-linking-covid-19-with-climate-crisis/

[20] Rob Toews, *Deepfakes are going to wreak havoc and we are not prepared*, FORBES (July 8, 2021, 03.00 PM) https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=2c84d82f7494

 [ISSN 2581-5369]