Page 951 - 968 **DOI:** *https://doij.org/10.10000/IJLSI.111027*

INTERNATIONAL JOURNAL OF LEGAL SCIENCE AND INNOVATION

[ISSN 2581-9453]

Volume 3 | Issue 4

2021

© 2021 International Journal of Legal Science and Innovation

Follow this and additional works at: https://www.ijlsi.com/

Under the aegis of VidhiAagaz – Inking Your Brain (https://www.vidhiaagaz.com)

This Article is brought to you for free and open access by the International Journal of Legal Science and Innovation at VidhiAagaz. It has been accepted for inclusion in International Journal of Legal Science and Innovation after due review.

In case of **any suggestion or complaint**, please contact **Gyan@vidhiaagaz.com.**

To submit your Manuscript for Publication at **International Journal of Legal Science and Innovation**, kindly email your Manuscript at **editor.ijlsi@gmail.com.**

Ethical and Legal Challenges of Deep Fakes - An Indian Perspective

SRISHTI DEY[1]

ABSTRACT

It is said that a camera cannot lie. However, in this digital era, it has become abundantly clear that it doesn't necessarily depict the truth. Increasingly sophisticated machine learning and artificial intelligence with inexpensive, easy to use and easily accessible video editing software are allowing more and more people to indulge in generating so-called deep fake videos, photos and audios.. These clips, which feature fabricated, altered and fake footage of people and things, are a growing concern in human society. Although political deep fakes are a new concern, pornographic deep fakes have been a problem for some time. These often purported to show a famous actress or model or any other woman, involved in a sex act but actually show the subject's face superimposed onto another woman's body who is actually involved in that act. This feature is called face-swapping and is known as the simplest method of creating a deep fake. There are numerous software applications that can be used for face-swapping, and the technology used is very advanced and is accessible. Deep fakes raise questions of personal reputation and control over one's image on the one hand and freedom of expression on the other. This will have a significant impact on user's privacy and security. Increasingly, governments around the world are reacting to these privacy evading applications for e.g., India banning TikTok and the USA investigating the privacy issues of TikTok and in the process of enacting laws to reduce the impact of deep fakes in the society. The study in this paper includes the ethical and legal implications surrounding the deep fake technology which also includes the study of several international legislations and analysing the position of India in tackling the crime of deepfake.

Keywords*: deepfake, artificial intelligence, data protection, privacy, ethical implications, legal implications, legislations, European Union, India*

I. INTRODUCTION

The world has seen ex-President of America, Mr. Barack Obama calling Donald Trump names (badmouthing), one of the big tech player, another video where, Mr. Mark Zuckerberg

[1] Author is a LLM Student at University of Petroleum and Energy Studies, Dehradun.

© 2021. International Journal of Legal Science and Innovation [ISSN 2581-9453]

declaring on having a "total control on billions of people's stolen data", and witnessing Jon Snow's moving apology for the unexpected and boring ending to Game of Thrones, all of these are products of deepfake. Deepfake uses a kind of artificial intelligence that is called deep learning to fabricate or alter images of fake evens, and hence the name as deepfake.

Synthetic media is a term where artificial intelligence algorithms are used for the modification, manipulation and artificial production of data by automated means. The deep fakes are the synthetic media used to create hoax videos, images and audios which are convincing enough to be believed as real. In simple words, deepfakes are any sort of videos or audios or images of any person that has been altered in a way that they appear to be of someone else and are often used with malicious intention. These clips, which feature fabricated, altered and fake footage of people and things, are a growing concern in human society. Although political deep fakes are a new concern in the scene, pornographic deep fakes have been a serious concern recently.The act of digitally stitching one's face to another's body is called face-swapping and is known as the simplest method of creating a deep fake. There are many software applications and websites that provide a platform for face-swapping, and the technology used is very advanced and is accessible. Deep fakes raise questions of individual reputation and control over one's privacy on the one hand and freedom of expression on the other. This will have a significant impact on user's privacy and security. European Union and several Governments around the world can be seen taking cognizance to analyse the harm it possesses and regulate the same.

II. A TECHNICAL OVERVIEW

A deepfake technology works with two machine learning models. One of those creates hoaxes from a set of datas of the available sample videos or images, where it processes how a face would blink, grin or smile and all other ways a face can animate its emotions. While the other model of machine learning tries to detect if the available sample detected from an autoencoder indeed is a hoax or not, when the second model can no longer detect or compare if the available video or image of the targeted person is indeed a hoax, then the deepfake generated product is probably believable enough as well to human eyes. This technique is called a generative adversarial network (GAN). It shall reject inaccurate samples of images or videos resulting in providing chances for more number of attempts to be generated, and thus, the cycle continues until a perfect version of that person is created. In short, an artificial intelligence backed robot creates an image of a person's facial expressions and another one keeps on telling if the captured expression looks fake and they argue till the end results are

nearly perfect.[2]

So, what exactly is Generative Adversarial Network (GAN)?

A GAN is a pair of adversarial Neural networks (ANN), ANN are capable of learning from unstructured audio-video data available on the internet, which can be used to create deepfakes.[3]

So, basically one network, that is a *generator*, that has the ability to generate from a latent sample input. The other network is called the *distributor*, that judges whether an input data is authentic or not. The discriminator thereby scores the fake data from 0 - 1, where 1 being the data with high probability of realness and 0 holding the high probability of the input data being fake. The generator thus uses the score received from the discriminator to adjust the weights until the discriminator can no longer differentiate between the real image and the fake image that has been generated by the generator.

GAN does better when the set data it is supposed to work on is available in large numbers and somehow, that is the reason the deepfake footages tend to feature big names, such as famous politicians, famous actors and other showbiz celebrities. They tend to have many videos on online platforms that GAN can use to create realistic deepfakes.

Deepfakes can be created using another method, which is less common and is known as *Variational Autoencoders* (VAEs). Unlike GAN, these depend upon two different networks that work together. The encoder network uses dense yet smaller representation of the received input data and the decoder utilises this data as output to produce original data. The decoder can then be blended and adjusted to create an effect that is desired. For example, a "face-swap deepfake," is generally mapping a person's face image onto a celebrity's body, and can be generated by adding two Variational Autoencoders. The face here is encoded by the means of that face encoder and then decoded using the celebrity decoder, creating a recreation of the original video with a new face easily believable to human eyes as an original result.[4]

III. CHALLENGES OF DEEPFAKES

Deepfakes are a major threat to human society as a whole. Deepfakes are capable of

[2] What is deepfake technology? https://www.techslang.com/what-is-deepfake-technology/ (last visited on Dec 26, 2020)

[3] Hasam Khalid, Simon S. Woo, *OC-FakeDect: Classifying Deepfakes Using One-class Variational Autoencode* ,CVF, Rev. 1, 2 (2020) rhttps://openaccess.thecvf.com/content_CVPRW_2020/papers/w39/Khalid_OC-FakeDect_Classifying_Deepfakes_Using_One-Class_Variational_Autoencoder_CVPRW_2020_paper.pdf

[4] Raina Davis, Chris Wiggins, Joan Donovan, *Deepfakes,* SPRING 2020 SERIES, Rev, 1, (2020) https://www.belfercenter.org/sites/default/files/files/publication/Deepfakes_2.pdf

contributing towards creating fake news to threaten national security by interfering in elections by disseminating fake propagandas on online platforms. The challenges that this technology possesses are really threatening.

Before categorising the challenges, first lets track down who produces deepfakes?[5] :

1) *Deepfake hobbyists* - Deepfake hobbyists communities are difficult to track down, they generally are up with swapping up faces of any normal person or celebrities' on bodies of porn stars', followed by making videoslike politicians say funny things, destroying one's marriage with an unreal sex video of either of the spouses, or derange an election by releasing a fake or unreal video or audio recording of one of the candidates prior to days before voting starts.[6] These hobbyists tend to see this kind of artificial intelligence generated videos as a new form of tech-art and use them to contribute towards online humor and look towards the development of such technology for solving out intellectual puzzles instead of using it to trick or threaten people. Whereas, some of them use it for their concrete personal benefits, such as for raising social awareness regarding the deepfake technology and the threats it possesses, whereas, they use it as art in order to get deepfake related work for several music videos, video games, advertisements or tv shows.

2) *Political players* - Political players, here includes the candidates, hackers, terrorists, and foreign states can use deepfakes in spreading fake political information, broadcast fake campaigns to manipulate public opinion and weaken the confidence of people in their country's institutions and its democracy.

3) *Fraudsters* - Fraudsters already have their hand in this threatening technology where artificial intelligence technology is used for the purpose of stock manipulation and other such financial crimes. They are already using AI generated fake audios to impersonate bank executives on the phone asking for bank card details, OTPs related to bank accounts and sudden cash transfers. In near future, artificial super intelligence would also be capable of faking live video calls and causing more damage to human lives.

4)*Entertainment Companies* - Deepfake technologies are used by several game developers to give face to the game characters, followed by using this technology in several music videos and movie scenes. These are used with the sole purpose of encouraging and showcasing the art of movie making.

[5]Westerlund, Mika. . *The Emergence of Deepfake Technology: A Review*. Technology Innovation Management Review. 9. 39-52, (2019) https://www.researchgate.net/publication/337644519_The_Emergence_of_Deepfake_Technology_A_Review.

[6]JM Porup, *How and why deepfake videos work — and what is at risk*, CSO INDIA, (10 April, 2019 15:30 PM) ,https://www.csoonline.com/article/3293002/deepfake-videos-how-and-why-they-work.html.

The deepfake technology is hereby backed by several challenges, which are further categorised in to - a) *Ethical challenges* b) *legal challenges*

A) Ethical Challenges

This technology gained popularity in the last few years. In December 2017, a social media user named as pseudonym 'deepfakes' showed the world how stitching faces digitally and using them maliciously is possible with the help of artificial intelligence techniques based on neural networks. Deepfakes then slowly gained popularity through funny and strange videos of famous showbiz celebrities and political figures on social media applications which were hard to believe to be unreal. Deepfakes have been recently receiving condemnation from around the globe for using of the technology for generating fake celebrity sex tapes, fake videos of politicians, financial frauds and revenge porn.

Some of the ethical challenges that this technology possesses can be categorised under Political challenges and Social challenges -

Political Challenges -

1) *Political misinformation* - World has witnessed several attempts of fake videos having a politician or public officer speaking his mind out and the video getting viral on social media platforms, with a malicious intent which had the capability to cause distress and spread fake information like fire in the land among the people. Fake political content created by using deepfake technology is a danger to society. Social media platforms like Facebook have been in constant pressure to remove the deepfake content from its platform. For example, a fake video of Obama badmouthing Donald Trump in 2018.[7]

2) *Political - social satire* - Sometimes a deepfake video is made having a politician's face and body but with different and a fake speech, the fake speech being very lite and funny. In that case the intent of the content is not to spread fake information but to elevate the social message behind the satire. Social media platforms like Facebook can be seen having to make attempts to distinguish between the deepfakes spreading fake information and deepfakes made for satire.

3) *Deep Fake news* - The journalism industry can be seen as a sufferer here due to its inability to saturate the fake and real content before broadcasting it to its viewers. The traditional fake news pose a lesser threat than the deepfakes do because they are harder to be

[7] Raina Davis, Chris Wiggins, Joan Donovan, Deepfakes, SPRING 2020 SERIES, Rev, 1, (2020) https://www.belfercenter.org/sites/default/files/files/publication/Deepfakes_2.pdf .

detected and people believe what they see as real. The technology is capable of producing seemingly legitimate news videos that place the reputation of the news agency.

Nowadays, a race to be the first one to provide the news to its viewers and to access the video footage shot by a witness of an incident can provide a competitive advantage to a news media agency and hence, to be the top one in the race they often miss out to verify if the footage is real or fake.[8] Wrongly attributed video footages of protest videos, accident videos, fake protest speeches and etc with wrong caption to suggest it happened somewhere else and shall raise concern somewhere else. For example, Reuters in New Zealand during the Christchurch mass shooting came across a viral video on the internet that claimed to show the moment where the suspect was being encountered by concerned security officials. However it was further discovered that the footage was actually from the USA and the suspect of the Christchurch mass shooting was not yet killed.[9]

4) *National Security* - Times are gone when the wars were fought with weapons and soldiers were deployed on war zones, followed by destruction of life and property. In this era of evolving artificial intelligence, wars are being fought on cyber space and with technology. Foreign interference with the elections and using this technology to spread fake political propaganda and disrupting election campaigns simply by releasing videos that go viral, and are done merely by putting new words in mouth of someone who's in powerful public position with intent to cause riots, violence, unrest, doubt and distress among the voters is a powerful weapon in today's fake information war.

Social Challenges

1) *Non-consensual and revenge porn* - The dark side of deepfakes, namely non-consensual and revenge porn, this technology enables the use of faces that are available online to be used in pornographic content without their consent. Here the famous celebrities are often the victims of non consensual porn that are created using the deepfake technology. Revenge porn using deepfake has the potential to violate victims' right to her own images and privacy, which is indeed a major concern.

2) *Blackmail and Extortion* - Deepfake pornographic videos of celebrities and celebrities falling prey to these videos are often heard, but these types of videos are now being used against ordinary men, women and children too. The deepfake porn videos consisting of

[8] Westerlund, Mika. The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review. 9. 39-52, (2019). https://www.researchgate.net/publication/337644519_The_Emergence_of_Deepfake_Technology_A_Review
[9] ibid.

ordinary faces are being posted just with an intention to shame, defame and humiliate the victim.[10]

3) *Financial Fraud and Cybercrime*: Here, the perpetrators often seemed to target CEOs using this deepfake technology, where it can make the victim say anything that he or she never said. Criminals could release a fabricated video that depicts a CEO falsely making comments that could affect the stock price of the company to fall while the perpetrators benefit from the short sales. This is a growing concern in financial fraud and white collar cybercrime in near future.[11]

4) *Human sounding synthetic voice* - Reports suggest that Google's is up with its efforts to develop voice assistant features that would be capable of mimicking human voice and with its addition to make and receive calls. Although voice assistants like Cortana, siri and Alexa are increasingly updating and are sounding more realistic.[12] A synthetic voice that sounds like a human raises several ethical concerns. As this technology can produce human-like voices, the probability is this synthetic voice technology is used to extort money from people by blackmailing them.

5) *Gender bias targets* - As majority of the victims of deepfake contents are women and they often fall prey for publishing non consensual pornographic content or any other kind of deepfake content with an intent of causing harm to her identity or shame. These social media platforms should uphold their respective platforms to be a safe place of entertainment for all. Though "*involuntary synthetic pornographic imagery*" was put to Google's ban list but still it is not enough to put a fullstop over its creation and its dissemination over cyberspace. The creation of deepfakes are generally done with an intention to threaten, blackmail, extort, cause harm to her individual identity, revenge porn and to silence women.

Once these altered and fabricated videos are shared online, it gets impossible to remove that from the social media platforms (internet). Often such kinds of contents go viral minutes after sharing and are shared, downloaded and uploaded multiple times. This gets impossible for the respective authorities to bring the content down and trace the IP address of the person who shared it. This technology has paved new ways of abusing and humiliating women.

[10] Francis Navarro, *Deepfake porn videos are now being used to publicly harass ordinary people*, KIM KOMANDO.COM, (Jan 01, 2019, 14:30 PM) https://www.komando.com/security-privacy/deepfake-porn-videos-are-now-being-used-to-publicly-harass-ordinary-people/526877/

[11] John Bateman, *Get Ready for Deepfakes to be Used in Financial Scams*, CARNEGIE FORUM FOR INTERNATIONAL PEACE, (Aug 10, 2020) https://carnegieendowment.org/2020/08/10/get-ready-for-deepfakes-to-be-used-in-financial-scams-pub-82469

[12] Natasha Lomas, "*Duplex shows Google failing at ethical and creative AI design*", TECH CRUNCH, (May 10, 2018, 1:57 AM). https://techcrunch.com/2018/05/10/duplex-shows-google-failing-at-ethical-and-creative-ai-design/

6) *Affecting market* - For defaming a product, tool, services, individual or a brand, deepfake technology can be easily used. This seems tough because a legal action or a suit can be filed only against a legal person. There are often fabricated online contents that can term certain products as harmful or poisonous made using the technology of deepfake. Mostly done by the rival companies to outpass the market competition. Most of the time, it is not possible to trace down the source of the content or verify the owner of the fake profile. Even if the source is traced down the act can be cleared by an alibi, it would most probably become too late and that might already have hampered the reputation.

7) *Tracing the source* - It is often seen that once these altered and fabricated videographic content or still images are shared online, it gets impossible to remove that from the social media platforms (internet). Often such kinds of contents go viral minutes after sharing and are shared, downloaded and uploaded multiple times. This gets impossible for the respective authorities to bring the content down from online platforms and trace the IP address of the person who created and first shared or uploaded it.

B) Legal Challenges

1) *Manipulation of Evidence* - Any evidence which is in the form of Audio - Visual form and is to be presented in the court has high chances of being altered with the use of deepfake technology, something that would be a roadblock to seeking truth and justice.[13] It poses a challenge on our institutions' role in redefining the lines of evidence and truth in the process of providing justice in the near future.

2) *Liar's dividend* - A new consequence is on the run, that is liar's dividend due to the recent trend of media fact checking to address fake information. This phenomena includes debunking fake information not only provides it a longer lifetime but also actually legitimizes its existence and the debate over its accuracy. While thinking about mitigation strategies the effective role of liar's dividend is also needed to be considered.[14]Related to rules of evidence and truth, in short, the liar's dividend is the ability for deepfakes to sow enough doubt in public in the Audio Visual content,in large that people start claiming the real one as a deepfake content and fake one as the real content.

[13] Britt Paris, Joan Donovan, *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence*, DATA SOCIETY, (2019). https://datasociety.net/wp-content/uploads/2019/09/DataSociety_Deepfakes_Cheap_Fakes.pdf .

[14] Paul Chadwickl. *The Liar's Dividend, and Other Challenges of Deep-Fake News*, THE GUARDIAN. GUARDIAN NEWS AND MEDIA (July 22, 2018, 19:00 BST). https://www.theguardian.com/commentisfree/2018/jul/22/deep-fake-news-donald-trump-vladimir-putin.

3) *Consumer Protection* - The deepfake generated social media filters raise privacy concerns over how the social media apps exploit user data in some instances.The viral Chinese Zao app, which allowed its users to swap their faces in their favourite movies and with the character they wanted to, by uploading their personal pictures. The terms of service of the app provided 'perpetual and transferable rights to the data uploaded', which raised major privacy concerns causing WeChat to ban the face swapping app Zao generated content.[15]

4) *Privacy* - The concern over the right to privacy being in danger due to the content being published using deepfakes because that covers non consensual manipulation, embellishment and distortion, as well as duplicitous uses of non-manipulated and original videographic content or still images for illustrative purposes.[16] The debate of privacy here arises over the use of celebrity faces as well as faces of ordinary people in non consensual pornography content. Sometimes it is often argued that an audio visual content shared publicly becomes private when manipulated.[17]

5) *Delayed Content Detection*: In accordance with the theory of liar's dividend the detection of deepfake shall always trail behind the production of deepfakes. This simply means that by the time the deepfake content is detected and is removed, it has chances of being consumed by the majority of the population. Policymakers must consider how to be pro-active in acting towards minimizing the possible harm that can be caused by circulation of a deepfake generated.

6) *Political and personal freedom* - Despiteful bots that are capable of producing fake news and social media content faster than any human writer, are behind the chaos caused in online media platforms.The role of deepfakes in creating and spreading wrong information and fake news challenges the main concept of fair and free elections. That creates a threat against violating the right to political participation and personal freedom.

Just like how people can easily use AI-powered deepfake technology to encourage the spread of fake information or are able to influence political public debate, they can easily use it to create and propagate and spread fake content designed to incite war or any kind of

[15] Grace Shao and Evelyn Chen, *The Chinese face-swapping app that went viral is taking the danger of 'deepfake' to the masses,* CNBC, (, Jan 17 20202:50 AM EST) https://www.cnbc.com/2019/09/04/chinese-face-swapping-app-zao-takes-dangers-of-deepfake-to-the-masses.html.

[16] David Fink, Sarah Diamond, *Deepfakes: 2020 and Beyond*, LAW.COM, (Sep 03, 2020 at 03:28 PM) https://store.law.com/Registration/Login.aspx?mode=silent&source=https%3A%2F%2Fwww.law.com%2Ftherecorder%2F2020%2F09%2F03%2Fdeepfakes-2020-and-beyond%2F .

[17] Clare McGlynn, Erika Rackley & Ruth Houghton, *Beyond 'Revenge Porn': The Continuum of Image-Based Sexual Abuse*, Fem Leg Stud 25. Rev.25, 46 (2017). https://doi.org/10.1007/s10691-017-9343-2

violence that can affect human life as well as property with their malicious and evil intentions.[18]

7) *Defamation* - A deepfake content can have anything, which may be related to an altered or fabricated videographic content or a still image. For an example, a video of an individual speaking some private information about another individual and that information tends to be true or an individual with public importance saying things in a video which he is not supposed to and the creator of the content just wanted to publish the content for fun or any other reason. In the first case, the victim for defamation here is the one who's private information is out, whereas, in the second case the individual in the video is defamed as he has been presented in a way that harms and derogates his image and hence is the victim. It is a technology that if not regulated precisely can cause harm to individuals of any social strata.

8) *Targeting women* - As majority of the victims of deepfake contents are women and they often fall prey for publishing non consensual pornographic content or any other kind of deepfake content with an intent of causing harm to her identity or shame. These social media platforms should uphold their respective platforms to be a safe place of entertainment for all. Though "*involuntary synthetic pornographic imagery*" was put to Google's ban list but still it is not enough to put a fullstop over its creation and its dissemination over cyberspace. The creation of deepfakes are generally done with an intention to threaten, blackmail, extort, cause harm to her individual identity, revenge porn and to silence women. The challenge on the policy makers are not only limited to data protection and privacy but also to ensure women safety with stricter laws.

IV. POSITION OF INDIA IN TACKLING CRIMES RELATED TO DEEPFAKE

India too has witnessed recent incidents related to deepfake revenge porn and deepfake technology being used in political campaigns. One was an incident in October 2019, where a man in Mumbai was arrested for an act of making a revenge deepfake porn video of his girlfriend just to threaten her.[19] Followed by in early months of 2020, in February two videos of Manoj Tiwari were released by BJP, where there was only one video but in two languages with an intention to reach two different linguistic voters.[20]

[18] *Human rights in the age of Artificial Intelligence,* https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf. Cal. Elec. Code § 20010(a) (2020)

[19] Parth Tyagi and Achyutam Bhatnagar, *Deepfakes and the Indian legal landscape*, INFORM.ORG ,(July 03, 2020) https://inforrm.org/2020/07/03/deepfakes-and-the-indian-legal-landscape-parth-tyagi-and-achyutam-bhatnagar

[20] Nilesh Christopher, *We've Just Seen the First Use of Deepfakes in an Indian Election Campaign*, VICE.COM (Feb 02, 2020) https://www.vice.com/en_in/article/jgedjb/the-first-use-of-deepfakes-in-indian-election-by-bjp. ,

The achievable solution right now to tackle the upcoming threat of this technology is to combine the technology and the legislation. The recent trend of using this deepfake technology in making fake pornographic videos and political campaigns do raise several questions over the concerns related to privacy, identity theft and as well as the reality and authenticity of elections and the content available in social media platforms.

There is no law that directly deals with deep fakes. There can be several causes of action already existing in our existing laws and can be extended as well to cover the deepfake crimes. Some of the provisions are as follows -

1) Defamation:

A person can be held liable for the act of defamation under the criminal as well as the civil law in India.[21] Also cyber defamation was earlier addressed under Section 66A of IT[22] Act. In civil law, defamation is punishable under the law of torts, where if the act is instituted and the act of defamation is found to be committed, damages shall be payable to the defamed person.[23] whereas, under the Indian criminal law, in section 499 of IPC, 1860[24] defamation is bailable, non-cognizable offence and compoundable offence and includes the publishing of some information that tends to cause some sort of harm to the reputation of the person. Punishment for the same has been provided under the Section 500 of IPC, 1860[25] which provides a sentence of imprisonment for up to two years or fine or both of these two.These laws are still not mature enough to deal with the various forms of existing deepfakes. Earlier cyber law too dealt with cyber defamation that was enshrined under Section 66A of IT[26] Act. The provision strictly covered any offensive information being sent by any computer source, with the intent of causing obstruction, insult, injury, hatred, criminal intimidation or ill will. This provision no longer exists in the IT Act, and hence was struck down by the Supreme Court in the case of Shreya Singhal v. Union of India.[27]

2) Right to Privacy:

In the case of *Justice K. S. Puttaswamy v. Union of India*,[28] The nine judge bench recognised

[21] T. Pradeep & Aswasthy Rajan, *A Critical Study on Cyber Defamation and Liability of ISPS*, 119(17) International Journal of Pure and Applied Mathematics, 1717, 1719 (2018)https://acadpubl.eu/hub/2018-119-17/2/139.pdf
[22] Information & technology, 2000, Section 66A.
[23]Rashmi Senthilkumar, *Defamation law in India*, LEGALSERVICEINDIA.COM (Jan 05, 2021, 18:10 PM) http://www.legalserviceindia.com/legal/article-2224-defamation-law-in-india.html
[24] Indian Penal Code, 1860, Section 499
[25] Indian Penal Code, 1860, Section 500
[26] Information & Technology Act, 2000, Section 66 A.
[27] A.I.R. 2015 S.C. 1523.
[28] (2017) 10 S.C.C. 1.

that a fundamental right to privacy is a fundamental right that is being protected under Part III of the Constitution of India, focused on the individual's right against the State and non-state actors for violations of their informational privacy, that recognizes an individual's control over his personal and digital privacy. This judgement was given in response to the argument of the AGI, where he considered that the Constitution of India does not include fundamental right to privacy with its connection to the legal challenge to Aadhaar Card (India's national identity project). Hence, using private or personal information like images, video clips in creating non consensual deepfake content of an individual shall institute violation of fundamental right to privacy.

Further, *Section 66E of IT Act, 2000* provides punishment for the violation of fundamental right to privacy if the accused person clicks or captures or publishes or transmits an image of a private area of any person without that person's express or implied consent, and does that knowingly or intentionally shall face imprisonment till three years or fine not exceeding the amount of two lakh rupees, or with both.

3) Offences Related to Computers

The misuse of deepfake technology and the contents related to that on the web are nonetheless offences to be covered under IT Act, 2000 as it is also a computer related offence. The publication of obscene data in an electronic form is duly punishable under Section *67 of the IT Act, 2000*.[29] *Section 67A of the IT Act,2000*, which also includes punishment for publishing content containing sexually explicit visuals in an electronic form.[30] Followed by, if the published material depicts children in sexually explicit form in an electronic platform shall be punishable under the Section 6*7B of IT Act, 2000*. If the deepfake content involves using any kind of unique identification feature, such as electronic passwords of an individual in a fraudulent manner, the accused person shall be punishable for the offence of, under the provided *Section 66C of IT Act, 2000*.[31]Furthermore, *66D of the IT Act, 2000* punishes for cheating by personation by using any computer resource.[32]

The Central Government, however, possesses the power to direct the intermediary to block any such deepfake content, if it finds necessary to do so, in the interest of protecting the sovereignty and integrity of India, national security, retaining friendly relations with the

[29] Information and Technology Act, 2000, Section 67.
[30] Information and Technology Act, 2000, Section 67A.
[31] Information & Technology Act, 2000, Section 66C
[32] Information & Technology Act, 2000, Section 66D,

foreign states or to maintain public peace and order.[33]

4) Copyright Infringement

Sometimes deepfake contents include altered versions of the sound and visual effects from a music video or a movie, which might be a copyrighted work. Section 14 of the Copyright Act, 1957 provides that the owner of that cinematographed music video or movie possesses exclusive right to license for making another copy of that film, including any picture or photograph of any image or any sound embodying it.[34] In the case of *Amarnath Sehgal v. Union of India*[35], Delhi High Court, the moral right of the author was recognised. The author can claim damages for the act of mutilation, distortion or any sort of modification that would be prejudicial to his honour and responsible for causing a violation of his moral right over his creation.[36] The Copyright owner is liable to receive civil remedies by way of injunction, damages and otherwise may be conferred by law for infringement of the moral right over his licensed work.[37] Furthermore, if any person who deliberately abets the violation of the copyrighted work or any other rights conferred to the copyright owner under the ambit of the Act shall be punishable with imprisonment that may extend till three years and fine with which shall extend to the amount of two lakh rupees.[38] But these remedies might not work for the victim of a deepfake content, because generally it is seen that copyright is owned by the producers of the movies, not the actors, who hold up the risks of being a target and same applies for the pictures and photographs, the copyright would be owned by the photographer not by the person in the photograph. So, the remedies provided under this act might not be advantageous for the actual victim or the target of the deepfake content.

5) Other Criminal Offences

The deepfake content can also be put under the radar of Section 468 of IPC, which defines the act of forgery, as the deepfake videos are generally the forged or copied versions of the original work and shall be liable to constitute the offense of forgery and which is created with an intent to harm the reputation or image of any party knowingly shall be punished with a sentence of imprisonment of either for aterm extended till three years and also shall be liable to fine.[39] Section 124 of IPC, 1860 applies for the punishment of an act, where a deepfake content is liable to spread hatred or contempt or excite disaffection towards the Government

[33] Information and Technology Act, 2000.Section 69A,
[34] The Copyright Act, 1957.Section 14(d) and Section 14(e),
[35] 117 (2005) DLT 717.
[36] Copyright Act, 1957.Section 57,
[37] Copyright Act, 1957.Section 55,
[38] Section 63, Copyright Act, 1957.
[39]Indian Penal Code, 1860, Section 469

of India. It shall be liable to punishment under the offence of sedition.[40]

Moreover Section 506 of IPC provides punishment for committing an offence in which there is a use of videos or images which threatens or intimidates any person or any of his property or his reputation.[41] Followed by the deepfake contents that tends to provoke breach of public peace and order[42], promoting communal outrage, promoting enmity between two religious or linguistic groups on the ground of caste, religion, race, language, place of birth or malicious acts to hurt religious feelings by insulting religious beliefs.[43]

6) Data Protection under The Personal Data Protection Bill, 2019 (PDP Bill) -

The legislature has introduced this bill with its aim to protect the privacy and data of an individual, so that by no means a personal data and privacy can be invaded without the consent of the data holder for processing of the personal data. The personal data includes one's image. The remedy for the violation of the rights protected under the act is right to be forgotten, which can be claimed against data fiduciary. The Bill is still not an Act yet. Moreover the right to be forgotten has already been recognised by some of the High Courts, such as Delhi, Kerala and Karnataka.[44]

V. RECOMMENDATIONS

1. **Laws and legislations** - one of the means against tackling deepfakes are strong laws and regulations. The existing state laws against defamation, identity fraud, data theft, privacy, copyright etc can be made more strong to deal with the crimes related to deepfakes under its legal ambit. Imposing strong punishment and heavy fines over creating a deepfake content might act as a brake towards the speed that this technology is evolving and the threat that it is posing in society. The regulators also frame laws that keep a thin line distinguishing the deepfake content created for expressing self under the right of freedom of expression, whereas another being violates one's right to privacy.

2. **Consumer protection and privacy** - The social media apps and service providers can develop and use techniques to alert users on viewing deepfake contents on their platform. There should be a limit to which the privacy policies of the service providers don't end up

[40] Indian Penal Code 1860, Section 128
[41] Indian Penal Code 1860, Section 506.
[42] Indian Penal Code, 1860.Section 504.
[43] Indian Penal Code 1860.Section 295A.
[44] Parth Tyagi and Achyutam Bhatnagar, *Deepfakes and the Indian legal landscape*, INFORM.ORG ,(July 03, 2020) https://inforrm.org/2020/07/03/deepfakes-and-the-indian-legal-landscape-parth-tyagi-and-achyutam-bhatnagar/

exploiting the data and privacy of its users through its terms and condition clauses. There should be efforts made to create secure digital infrastructure.

3. **Education and awareness** - Education and awareness is necessary for combatting deepfake contents. Public should be aware and should be capable of not falling for deepfake contents and possible threats that these AI generated technology possesses. Furthermore, educating the youth on not sharing or promoting any fabricated content online. Public must also be aware of how to identify a deep fake content that would be available on online media platforms.

Some of the subtle indicators that can help identifying a fabricated or altered content are such as[45] -

a. The eyes of the fabricated face of a deepfake content might not blink humanly or may look unnatural. The blinks would not match a normal human blinking cycle.

b. It is often seen that the lip synching of the fabricated face be either off or not in sync as per the audio. In that case it is quite an indicator that the audio video is a mismatch.

c. With an irregular lip sync, the movement of teeth might too look unnatural.

d. There is an unnatural movement of facial muscles, followed by patchy skin texture.

e. In case of an individual having long hairs, the movement of the hairs would look unnatural.

4. **National Security and democracy** - The policy makers should be aware of the threat that these deepfake contents possess and how it can hamper national security. There is a need for a strong technical wall to be built so that it doesn't allow any deepfake content sow the seed of hatred, riots and other kinds if violence within the state.

5. **Eligible justice systems** - The legal officers, attorneys and the police officials should be trained to handle the cases related to deepfake crimes, including the necessary techniques required to store and analyse the evidence, such as source of the content, intention of dissemination of the content etc.

[45] David Fink and Sarah Diamond, *Deepfakes: 2020 and Beyond*, LAW.COM (Sep 03, 2020 at 03:28 PM) (https://www.law.com/therecorder/2020/09/03/deepfakes-2020-and-beyond/

6. **Moral obligation** - The moral obligation of the creators and distributors of deepfakes shall ensure that they ethically use and employ such kind of synthetic media. The companies that provide tooling and cloud computing, such as Amazon, Google and Microsoft to create deepfakes too have a moral obligation to put some checks and balances on its technical policies.[46]

7. **Responsibility of Social media platforms** - The famous social media platforms such as Facebook, Twitter, TikTok, Instagram, Snapchat and etc, which provides its users to upload and share contents must have and own some ethical and social responsibility on banning or prohibiting the publishing of deepfake contents.[47] As majority of the victims of deepfake contents are women and they often fall prey for publishing non consensual pornographic content or any other kind of deepfake content with an intent of causing harm to her identity or shame. These social media platforms should uphold their respective platforms to be a safe place of entertainment for all.

8. **Marking the fabricated content** - Most often the deepfake content shared online can often skip people's mind for the thought of it being real but truth seems otherwise. One approach that can be used is to add disclaimer watermarks in the content, sch as "fake content" or any appropriate watermark while it firstly gets uploaded on any online platform by use of video fingerprinting. Which can be of some help in solving the problem to a certain extent.

9. **Tracing the origin** - The approach that can be used here to trace the origin of the generated deepfake content is to direct a reverse image search to find the similar content that appeared priorly and the recent research shall hereby suggest the distributed technological solutions such as the Blockchains, which can undoubtedly help in protecting the society from the trending fake news and deepfakes.[48]

10. **Theory of *Authenticated alibi service***: This theory named as authenticated alibi service was put forward by Danielle C. and Robert Cheshney, which suggests that they could eventually create digital life-logs that are capable to track down the individual so that they eventually disapprove the claim being portrayed by the altered or fabricated content online.

[46] Ashish Jaiman, *Debating the ethics of deepfakes*,ORFONLINE.ORG, (Jan 07, 2021, 23:09 PM) https://www.orfonline.org/expert-speak/debating-the-ethics-of-deepfakes/#_edn9

[47] Mira Lane, "Responsible Innovation: The Next Wave of Design Thinking",MEDIUM.COM, (May 19, 2020) https://medium.com/microsoft-design/responsible-innovation-the-next-wave-of-design-thinking-86bc9e9a8ae8

[48] Minna, Deepfakes: An Unknown and Uncharted Legal Landscape, TOWARDS DATASCIENCE.COM (Jul 17, 2019) https://towardsdatascience.com/deepfakes-an-unknown-and-uncharted-legal-landscape-faec3b092eaf

This solution helps in rejecting potential fake information and helping out protecting one's privacy.[49]

VI. CONCLUSION

Technology is developing day-by-day. Everyday there is some new addition in the aspect of technology. However, law does not develop at such pace and the present laws in India and several other countries don't have any legislation primarily regulating the deepfakes. The existing laws might not be sufficient to address the deepfake issues using technological algorithms. There might arise some issues on regulating deepfakes, such as:

- Recognition and identification of deepfakes in real time.
- Attribution can be proved and culprits be punished.
- Acknowledging the gap in recognising if the content published supports the notion of freedom of speech or violates one's right to privacy.
- Ensuring that benefits for the victims is not outweighed while going through these lawsuits.
- The effect of the inherent rhythm of the courts, that needs to be restored in accordance with the current need of controlling and mitigating the effects of deepfakes.
- If the police departments are well equipped and trained to conduct investigations relating to deepfake contents.
- Technical knowledge that the legal attorneys must possess to conduct these types of criminal accusations.
- Removal of the deepfake content from the internet as early as possible.

These issues are long debated in the context of cybersecurity. But we as a society, too, have a moral obligation to help curb the spread of non-consensus malicious content. Educating ourselves and spreading awareness regarding the manipulations and the harm it can cause. Youth should be taught regarding the consequences of creating, uploading, downloading or sharing of fabricated content online. The regulators should look forward to adopting new methods in regulating the issues related to deepfakes, so that the source of the content is identified and blocked accordingly. It is often seen that the main defence used against the fake content is that an individual on one hand has the freedom of speech and expression granted under Article 19 of the Constitution of India. The thing that is to be considered is that

[49] MINNA, *supra note 67*, at 68.

our freedom of expression ends where one's right to privacy begins. Our duty here is to understand that our actions and freedom does not tend to hamper any other individual's enjoyment of rights. The right to withholding an assent is a right guaranteed under Article 19 of Indian Constitution to every individual but the same can't be used to justify the creation and dissemination of fabricated or altered videographic content/still image that has the capability to manipulate people's thought process regarding the subject of the content. Hence, to combat this, the regulators and citizens should do their duty towards the welfare of the society.

*****

© 2021. International Journal of Legal Science and Innovation [ISSN 2581-9453]