Deepfakes: A Challenge for Women Security and Privacy

Dr. Zubair Ahmed Khan*

Ms. Asma Rizvi**

Introduction

A deepfake is technology that is entirely or partially generated or modified (video, audio, or otherwise). "Deepfake" is defined as the form of Artificial Intelligence (AI) and Machine Learning (ML) technologies to analyze and generate the altered digital media content, such as images and videos, to falsely portray a person as saying and doing something that is not actually, they do or say. SIGGRAPH Asia Conference held on 19th-20th November 2019 in Tokyo, Japan. Deepfake, as per International Conference on Computers and Applications in Security (ICCAS), New York City, USA is a type of media manipulation in which Artificial Intelligence (AI) and Machine Learning (ML) techniques are utilized to generate or alter digital media content, such as images and videos, to falsely portray a person as saying or doing something they did not. Deepfake technology is such kind of technology where tone, modulation and facial expression can be adjusted in a single frame and distinguished features of 2 or more individual can also be combined for the requirement having high quality.[1]

**Assistant Professor, University School of Law & Legal Studies (USLLS) ,Guru Gobind Singh Indraprastha University*

** Research Scholar, *University School of Law & Legal Studies (USLLS) ,Guru Gobind Singh Indraprastha University*.

[1] Ian J. Goodfellow et al., "Generative Adversarial Nets", 27TH ANNUAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NIPS 2014), MONTREAL, (Dec. 8, 2014), https://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

Deepfakes are a type of artificial media technology which can generate realistic-looking audio, video, or image content using a variety of techniques such as facial recognition, voice synthesis, and machine learning. The technology is being utilized in creating realistic-looking videos of people, often known as “deepfakes”, which have been used to maliciously target women, including creating videos and images that contain explicit content, or resemble them in certain way that it can damage their reputation. The potential harms to women posed by deepfakes are numerous and far-reaching. For example, deepfakes can be used to spread false and damaging information about women, either maliciously or inadvertently, and they are used to manipulate the public’s perception of women. Additionally, deepfakes are used to create fake nude images of women, or to blackmail them by threatening to release compromising images or videos. The legal implications of deepfakes for women’s security and privacy are also concerning. In the United States, for example, deepfakes is considered to be infringement of different legislation including Computer Fraud and Abuse Act, the Digital Millennium Copyright Act, and the Defamation Act, and state-level laws as well.

It is matter of concern that many cyber criminals know for abusing deepfakes technology for their own interest. The outright prohibition of this technology or any AI based technology is not an appropriate and remedial measures against increasing case of cyber-crime against women on internet. Thus, there is need for periodic assessment in the interest of digital governance and fix accountability in the internet regime.

Misuse of Deepfakes as technology

False videos were extremely uncommon and generally impossible to produce because of "the scarcity of advanced equipment, the huge demand of specialized knowledge of subject matter, and the difficult and time-consuming technique involved."[2] The reason is "the abundance of huge amount of training data and high-throughput computing capacity" and "the development of machine learning and computer vision algorithms that exempts from the for manual editing methods".[3] Deep fakes are now commonplace and easy to produce.[4] A deepfake video is created by utilizing a source material consisting of films or other visual representations of an individual, which is then processed to generate a modified video depicting the subject's face. This process is called as deep learning. The outcome of it is created through neural networks, which are "a type of Machine learning (ML) where a computer executes a task by assessing cases". The neural network is instructed to "automatically detect the face expression of the subject" before the updated video can be created[5], which allows the network to generate an updated video. Generative adversarial networks, often known as GANs, are the most up-to-date method for constructing deep fakes. This method "contains two neural network that are trained parallelly," according to one

[2] Yuezun Li & Siwei Lyu, "Exposing DeepFake Videos by Detecting Face Warping Artifacts", arXivLabs, CORNELL UNIVERSITY, (May 22, 2019), https://arxiv.org/pdf/1811.00656.pdf.

[3] *Ibid.*

[4] *Ibid.*

[5]Larry Hardesty, "Explained: Neural Networks", MIT NEWS (Apr. 14, 2017), http://news.mit.edu/2017/explained-neural-networks-deep-learning-0414.

definition.[6] After attempting to understand the statistical patterns present in a data, like a collection of images and videos, a neural network called the "actor" will subsequently produce convincing synthetic data based on those learned patterns. The other neural network, which is referred to as the "critic," distinguishes with the authentic and fictitious samples.[7] The final outcome is an iterative procedure wherein the feedback from the second neural network aids in the progressive enhancement of the first neural network's ability to generate deepfakes that exhibit a rising degree of realism.[8] They compete to outwit one another to confuse the art detector where one cannot tell the difference between real and fake objects.[9]

Deepfake audio clips are utilizing processes that are quite same. This is not a costly technology, and amateurs are finding it easier and easier to get their hands on it. Lyrebird AI in 2017, developed technology for "voice cloning" and circulated fake audio recordings, including one the then President of United States Mr. Donald Trump addressing the issue of sanctions on North Korea.[10] A claimed audio recording of President Trump saying, "The United States is considering banning all trade with any country doing business with North Korea," was used to demonstrate this method in the report. It was

[6] Will Knight, "The US Military Is Funding an Effort to Catch Deepfakes and Other AI Trickery", MIT TECH. REV. (May 23, 2018), https://www.technologyreview.com/2018/05/23/142770/the-us-military-is-funding-an-effort-to-catch-deepfakes-and-other-ai-trickery/.

[7] *Ibid.*

[8] Martin Giles, "The GAN father: The Man Who's Given Machines the Gift of Imagination", MIT TECH. REV. (Feb. 21, 2018), https://www.technologyreview.com/2018/02/21/145289/the-ganfather-the-man-whos-given-machines-the-gift-of-imagination/.

[9] *Ibid.*

[10] Lyrebird AI (@LyrebirdAi), TWITTER (Sept. 4, 2017, 2:41 AM) , https://twitter.com/LyrebirdAi/status/9045953269 2 174528.

determined that the technology behind Lyrebird AI is "wonderful," and that it would "improve with time."[11] Despite this, there is still some questions regarding to which it is successful. In another case, a corporation fell victim to a fraud involving counterfeit audio recordings. This incident highlights the challenges associated with verifying the validity of fake audio clips, which may be more complex compared to counterfeit video clips. Consequently, the presence of fake audio clips presents a greater risk.[12] The inventor of deepfakes, Hao Li, predicted in Sept. 19' that totally real deepfakes will be available to the general public within six to twelve months' time.[13] A single image and a neural network that has been trained on a large data set consisting of images and videos may now be used by Samsung's software, which is now open for sale, to generate a deepfake that seems to be incredibly lifelike.[14]

Machine Learning is frequently available via a variety of commercial services.[15] The capability to easily construct deepfakes technology is a handy

[11] James Vincent, "Lyrebird Claims It Can Recreate Any Voice Using Just One Minute of Sample Audio", VERGE (Apr. 24, 2017, 12:04 PM), https://www.theverge.com/2017 /4/24/15406882/ai-voice-synthesis-copy-human-speech-lyrebird.

[12] Catherine Stupp, "Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case", WALL ST. J. (Aug. 30, 2019, 12:52 PM), https://www.wsj.com /articles/fraudsters-use-ai-to-mimic-ceos-voicein-unusual-cybercrime-case-11567157402.

[13] Kevin Stankiewicz, 'Perfectly Real' Deepfakes Will Arrive in 6 Months to a Year, CNBC: TECH (Jan. 17, 2020, 2:51 AM), https://www.cnbc.com/2019/09/20/hao-li-perfectly-real-deepfakes-will-arrive-in-6-monthsto-a-year.html.

[14] Joan E. Solsman, "Samsung Deepfake AI Could Fabricate a Video of You from a Single Profile Pic", CNET (May 24, 2019, 7:00 AM), https://www.cnet.com/news/samsung-aideepfake-can-fabricate-a-video-of-you-from-a-single-photo-mona-lisa-cheapfakedumbfake/.

[15] Kashmir Hill & Jeremy White, "Designed to Deceive: Do These People Look Real to You?", N.Y.TIMES (Nov. 21, 2020),

one that they will spread to the general public very easily. This accessibility makes deepfakes makes it more dangerous because it can be used to steal private and sensitive details of individuals. The use of social media platforms and internet platforms as a means for the mass dissemination of generated videos or photos through deepfake is something that further complicates the situation. The rapid and cost-effective distribution of images, audio, and video on social media and online platforms, coupled with cognitive biases that lead individuals to share negative, novel, or belief-confirming information, is expected to contribute to the widespread spread of scandalous or harmful deepfake content. The combination of these two circumstances implies that this outcome is likely to occur.

Beneficial use of Deepfakes technology

Deepfake technology is being utilized in a wide variety of contexts, for example the film industry, social media platforms for different social and commercial purpose, entertainment sector, health sector,etc. The entertainment industry is getting different commercial advantage from Deepfake technology in diverse manner. For instance, to assist in the modification of digital voices for performer who died due to illness or any other sudden accident, or it can be used to update film material rather than reshooting it. Both of these applications save time and money. In the future, it will be easy for movie-makers for reproducing movie scene, videos with actors who might have passed away, create special effect, suitable facial edit as per requirement and something which is unanticipated in the entertainment industry. The Deepfake technology may help crucial role in getting

https://www.nytimes.com/interactive/2020/11/21/science/artificial-intelligence-fake-peoplefaces.html.

automated voice for dubbing in movies, documentaries of different language and other virtual media in the interest of diverse type of people conducive to regional preferences.

An educational advertisement featuring David Beckham in 2019 global campaign to discontinue use of malaria used technology to make him appear to speak other languages, both via his appearance and through his voice.[16] Deepfake technology, on the other hand, is able to translate speech during video conference calls while altering facial and mouth movements to improve eye contact and give the impression that everyone is speaking the same language. This enables language barriers to be overcome and allows everyone seems to be communicating in the same language. Hence, the development of websites that provide access to Deepfake-induced birthday videos of various celebrities wishing people happy birthday, is the direct violations of celebrity rights of celebrities who do not sell their rights to particular websites has grown.[17]

Deepfakes is a technology that enables more telepresence in virtual games and chats worlds, as well as assistants that sound and appear authentic and are incredibly intelligent, and digital doubles of real people. This helps to improve both interpersonal connections and the quality of engagement

[16] Omar Oakes,"Deepfake voice tech used for good in David Beckham malaria campaign", THE CAMPAIGN, Apr 9th 2019, https://www.campaignlive.co.uk/article/deepfake-voice-tech-used-good-david-beckham-malaria-campaign/1581378.

[17] Tring- website for wishes by celebrities, https://www.tring.co.in/tring-dhamaka/all?utm_source=Google-GenE-Search&utm_medium=googlecpc&utm_campaign=&utm_adgroup=&utm_term=get%20birthday%20wish%20from%20celebrity&gclid=Cj0KCQiAw8OeBhCeARIsAGxWtUy0T9bimNKUZMoCWM640Pg2GecBOEbsLq3eXON1pzJ-K0EUqIzHj0UaAslYEALw_wcB > https://wearenova.ai/birthday-video-maker/ (accessed on Mar 21st 2023).

happening online. In a similar vein, the technology may be put to beneficial use in the disciplines of social work and medicine. Deepfakes is a tool that assist people in coping up with the death of loved ones by digitally "resurrecting" a departed friend. This gives the bereaved loved one the opportunity to say their final goodbyes to the deceased friend. It also has the capability of digitally recreating a missing limb or assisting transgender persons in seeing themselves in their desired gender. The Deepfake technology can potentially let patients with Alzheimer's illness communicate with a younger version of themselves that they may remember.[18] Also, researchers are trying to use GANs to identify X-ray abnormalities, and their potential for speeding up the discovery of materials science and medicinal breakthroughs by enabling to create the virtual chemical compounds. Because it has the potential to drastically revolutionize e-commerce and advertising, companies are interested in the brand-applicable deepfake technology. For example, brands can show fashion outfits on a diverse range of models with differing skin tones, heights, and weights. These models can be supermodels who aren't really supermodels who have been hired by the brand. Deepfakes also make it possible to create highly personalized content by turning ordinary people into models. But apart from creation of artificial voice and innovative utility, there is apprehension of misuse and encouragement of digital clones.

[18]Esat Dedezade, "I-Remember: an app that helps people with Alzheimer's recognize faces using AI", MICROSOFT FEATURES, 17 March, 2020, https://news.microsoft.com/europe/features/i-remember-an-app-that-helps-people-with-alzheimers-recognize-faces-using-ai/.

Deepfakes and cyber-crime against women challenge to women security (Revenge pornography, Celebrity porn, fake porn, sextortion, face morphing, spoofing of voices)

Pornography had the first deepfake attack. 96% of deepfakes are pornographic, with over 135 million views on pornographic websites, according to sensity.ai. Deepfake pornography targets women alone but for the purpose of extorting money sometimes men are also easy victims of Deepfake pornography. Pornographic deepfakes may scare and damage and image and life of individuals. It reduces women to sex object, inflicting emotional suffering, financial loss, and job loss.[19]

Revenge Porn- Two Zimbabwean women recounted their experiences as victims of revenge porn on BBC's The She Word; one was disowned and consequently unable to complete her education, while the other lost her job.[20] After becoming the subject of image-based sexual assault, it is totally understandable for women to find it challenging to keep or acquire work, as stated by several. To add salt to injury, certain internet service providers where this form of online abuse occurs were slow to recognize the problem, and addressing it remains difficult. The violation of privacy, indecent representation of women are serious concerns associated with this type of heinous crime.

For revenge porn victims it is common that they suffer from anxiety or despair, PTSD, or substance addiction. Interestingly, a study emphasized that

[19] Ashish Jaiman, "The danger of deepfakes", THE HINDU, 1st Jan 2023, https://www.thehindu.com/sci-tech/technology/the-danger-of-deepfakes/article66327991.ece.

[20] "The She World", BBC WORLD SCIENCE TV, 19th Dec 2019, https://www.bbc.co.uk/programmes/p07xs7qs.

male victims of image-based sexual abuse report feeling less guilt and less self-blame than female victims in the same circumstance.[21] Without implying causation, male victims reported a larger proportion of favorable police reactions when they reported their case than female victims. To be examined, however, is whether the more unfavorable police response to women reporting abuse reflects a broader social outcome that the women could have done more to avoid violence, sometimes known as victim-blaming.

Celebrity Porn- Another form of deep fake sex material refers to face-swapped celebrity porn where famous photos are placed on the bodies of persons participating in sexual acts. Scarlett Johannsson and Gal Gadot are just two of the well-known actresses who have appeared in the earliest examples of deep fake sex material. Yet anyone could be a victim of having his or her face superimposed onto the bodies of porn stars engaged in sexual acts.[22]

Sextortion, also “eWhoring”- Sextortion is a relatively new combination of the words like "sex" and "extortion." In general, extortion happens when "one individual takes advantage of another against his or her will by threatening him or her with violence or injury of any kind"[23]. The harm can be physical (to them or their loved ones) or can target their property or reputation[24],

[21] Youtube, “Meet The Women Being Deep faked Into Porn by AI | Deepfake Porn: Could You Be Next?”, BBC THREE, 3 Nov. 2022, https://www.youtube.com/watch?v=Q-S-amtvcd8.

[22] K Melville, “The insidious rise of deepfake porn videos — and one woman who won't be silenced”, ABC NEWS, 30 Aug. 2019, 11:00AM, https://www.abc.net.au/news/2019-08-30/deepfake-revenge-porn-noelle-martin-story-of-image-based-abuse/11437774 (accessed 1 Mar. 2023).

[23] Forsyth, C. J., & Copes, H. (2014). Encyclopedia of social deviance. SAGE, pp 266.

[24] Konrad, K. A., & Skaperdas, S. (1998), ECONOMICA, 65, pp 461-477

typically involving blackmail[25] -the threat of revealing damaging secret information or ransom, where something of value is held until the victim fulfils a specified condition. The term sextortion has evolved to describe incidents of extortion in which a person threatens to distribute sexually graphic photos that has been collected, voluntarily or not. Sextortion can be particularly successful because it is not perceived as the technology of deepfake for creating morphed and fake explicit pictures of individual. It is essential for people to aware with deepfake technology and the possible dangers that might be posed and to take measures to safeguard themselves against those dangers.

Spoofing of voice- Mimicking and playing back altered audio aren't the only ways to fake a voice. Audio information is manipulated or generated using advanced machine learning and AI techniques in deepfakes. Deepfake speech synthesis is becoming better and better every day. Voice recognition is used by several government organizations for identification verification, and by many banks for wire transfers and online banking. On 30TH May 2023, Ministry of Home Affairs, India has claimed the government of India has disabled over 500 applications used for spoofing and fraud.[26] The applications were prohibited on Indian Cybercrime Coordination Centre's (I4C) recommendations for security concerns and action was done as per legislation. States have received an analytical report on the top 50 cyber-

[25]Lindgren, J., "The theory, history, and practice of the bribery-extortion distinction", UNIVERSITY OF PENNSYLVANIA LAW REVIEW, 141, pp 1695-1740. 1993, https://scholarship.law.upenn.edu/penn_law_review/vol141/iss5/6..

[26] Anamica Singh, "India blocks over 500 apps over spoofing and fraud fears", WION NEWS, Mar 30th 2023, https://www.wionews.com/technology/india-blocks-over-500-apps-over-spoofing-and-fraud-fears-amit-shah-576940 (accessed on Mar 21st 2023).

attacks' methods. The investigation includes AIIMS cyberattack.[27] The inquiry uncovered four IP addresses and previously undisclosed email addresses and phone numbers.

References obtained via cases all throughout the world

There have been many high-profile examples involving the exploitation of deepfake technology for harmful objectives, like the fabrication of non-consensual pornography and sextortion. These cases have received full media attention. The following are some examples: The technology is also advancing every second, as seen by a new software for Android that was only developed not too long ago and that can be used to generate nude photographs of women in a matter of minutes. Numerous instances of revenge pornography have been documented throughout India's history.[28] In 2014, a college student from the city of Udupi in the Indian state of Kerala was astounded to learn that her ex-boyfriend had uploaded many private photographs of her to the internet and shared them with others.[29] In yet another horrifying invasion of privacy and betrayal of trust by a separated

[27] Kaunain Sherriff M, "AIIMS cyber-attack: At least five servers infected, have data of 3-4 crore patients", THE INDIAN EXPRESS, (Nov 20th 2022), https://indianexpress.com/article/cities/delhi/aiims-cyber-attack-at-least-five-servers-infected-have-data-of-3-4-crore-patients-8297028/ (accessed on Mar 21st 2023).

[28] David K,"Deepfake App Creates Nude Images of Women in Seconds", THE VERGE, June 27th 2019, https://deepfake-nude-ai-app-women-deepnude-non-consensual-pornography (accessed on Mar 14th 2023).

[29]Haritha John, Nayantara N., "Google Cracks Down on Revenge Porn, Here's What Indians Should Know", THE NEWS MINUTE,(June 22nd 2019), https://www.thenewsminute.com/article/google-cracks-down-revenge-porn-heres-what-indians-should-know-31437 (accessed on April 16th 2020).

couple. The Deepfake app can produce nude pictures of women in a matter of seconds.

A man in Ahmadabad was accused of portraying his wife and sister-in-law as prostitutes on social platform by posting private chats and naked images of his wife and sister-in-law.[30] In a related incident that took place in the same location, an ex-husband was found to have emailed personal images of himself and his wife taken during their honeymoon to his sister-in-law and her spouse.[31] After the breakup of his relationship with his fiancée, a young man from West Bengal who was 23 years old shared a naked video with her.[32] The girl was the one who first recorded the video, but the male used it as a means to sexually exploit her and satisfy his own sexual desires. The appropriate legal action was taken against him by the Federal Bureau of Investigation (FBI), but the harm had already been done.

The fashion of revenge porn in Indian society, thousands of innocent women are used for the vile pleasure of vengeance, which is a problem that has to be addressed as soon as possible. However, deepfakes will make the problem far more widespread. It is simple to create fake pornographic movies and photographs in nude state and post them over social media. Hence, the morals and decency of society are under significant assault from the trend of

[30]TNN, "Separated Husband Posts Intimate Pictures", THE TIMES OF INDIA, (July 13th 2017) https://timesofindia.indiatimes.com/city/ahmedabad/separated-husband-posts-intimate pictures/articleshow/59568022.cms (accessed on April 15th 2020).

[31] *Supra.*

[32]Priya Pathak, "Revenge Porn : In a First, 5-Year Jail for Indian Man Who Shared Nude Video of Ex Girlfriend", INDIA TODAY,(Mar 12th 2018), https://www.indiatoday.in/technology/news/story/revenge-porn-in-a-first-5-year-jail-for-indian-man-who-shared-nude-video-of-ex-girlfriend-1187451-2018-03-12.

deepfakes. The possibility that there have been further cases as well, and critical that laws and procedures be in place to deal with this problem.

United States laws against Deepfakes

From past few years, many states in the United States of America have passed laws that put restrictions on the potentially harmful usage of deepfakes.[33] Yet, the right to free speech guaranteed under First Amendment places significant limitations on the use of this law. The courts will rule that this state-level legislation breaches the principles outlined in the Constitution. Under "Assembly Bill No. 730, California," which was approved by the Governor on October 03, 2019, the use of deepfakes in election materials was made illegal in California in 2019.[34] This was accomplished by making it illegal to manufacture or distribute "materially false" campaign materials within sixty days of the election, which was mandated by "Assembly Bill No. 730, California."[35] If a reasonable person would have a fundamentally different understanding or perspective of the content if the original, unmodified image were revealed, and then considered dishonest in the usage of those generated images. This standard will remain in place until 2023. Notably, media that constitutes satire or parody is exempt from the prohibition. So are news broadcasts that publish the images as a genuine news story, websites, and

[33] Penelope Thornton, "Deepfakes: An EU and U.S. perspective", GLOBAL MEDIA TECHNOLOGY AND COMMUNICATIONS QUATERLY (GMTCQ) SPRING SUMMER 2020, https://f.datasrvr.com/fr1/320/16758/1207330_-_GMCQ_-_Spring_2020_Deepfakes.pdf, pp 30-36.

[34] *Ibid.*

[35] Kari Paul, "California Makes 'Deepfake' Videos Illegal, But Law May Be Hard to Enforce", The Guardian (Oct. 7, 2019), https://www.theguardian.com/us-news/2019/oct/07/california-makes-deepfake-videos-illegal-but-law-may-be-hard-to-enforce (accessed on Mar 21st 2023).

regularly published periodicals, provided that the dissemination is accompanied by a clear acknowledgement, depending on the circumstances, that it is inaccurate or its authenticity is in question. Also exempt from the prohibition are media that constitutes satire or parody.

The California Act against deepfake, known as AB730, also includes a broad exception for broadcasting stations that are compensated to transmit materially false media. This exception applies regardless of whether the station issues a disclaimer indicating the lack of authenticity in the content it transmits. In 2019 an effort to regulate the usage of pornographic deepfake images, the state of California granted a specific right to civil damages to individuals who were depicted in sexually explicit materials without their consent. This right was granted to these individuals regardless of whether or not the individuals participated in the creation or development of the sexually explicit materials.

Under Virginia House Bill 2678[36], which was passed on March 18, 2019, the state of Virginia revised its statutes criminalizing the unlawful sharing of sexually explicit materials with malicious intent to include the dissemination of changed photos with the purpose to portray a real, identifiable person. These changes were made in response to the passage of the bill. Despite that New York is considering passing a bill agaisnt deepfakes, the state has already passed a legislature that would protect the digital likeness of individuals. As a component of its "Right to Privacy" statute, the publicity right in New York is enshrined in Article 5 of the New York Civil Rights Law. Sections 50 and

[36] Virginia Bans, "Deepfakes'and 'Deepnudes' Pornography", BBC News, July 2nd 2019), https://www.bbc.com/news/technology-48839758#:~:text=Virginia%20has%20become%20one%20of,videos%20without%20the%20victim's%20consent (accessed on Mar 21st 2023).

51 of the Act are the primary legal provisions that apply to procedures involving the publicity right.[37] The two sections 50 and 51 describe rights that are analogous to one another but provide different procedures for their enforcement. Contrary to the provision in Section 51, which provides for a private cause of action, Section 50 makes it a felony to violate the publicity right. Under Section 51, you have the right of the protection of your name, face, image, and voice.[38] Posthumous publicity is not a privilege that is recognized in New York. The portrait and image sections of the Act have been broadly construed by the courts to mean that they embrace "any recognizable likeness, not only an actual photograph." A plaintiff has the right to seek both an injunction to prevent the continued use of her identity and monetary compensation for the harm that was caused by the identity being exploited in the past. In most of the judgments, damages are intended to compensate for the plaintiff's emotional suffering.[39]

The United States Congress has passed legislation that makes it easier to collect data on deepfakes, which is currently debating new legislation that, if enacted, would make it mandatory for more research and reporting to be conducted on deepfake media and the technology that is used to create deepfakes. The Identifying Outputs of Generative Adversarial Networks (IOGAN) Act of 2019, which was approved by the House of Representatives in December 2019 and is currently being reviewed by the Senate, would mandate that the National Science Foundation and the National Institute of Standards and Technology fund research on "generative adversarial

[37] Mirone v. MacMillan, 894 F.2d 579, 585 (2d Cir. 1990).

[38] Burck v. Mars, Inc., 571 F. Supp. 2d 446, 451 (S.D.N.Y. 2008).

[39] Garis v. Uncut-RawTV, No. CV 06–5031, 2011 WL 4404035, at *3-4 (E.D.N.Y, July 5, 2011).

networks," which are software programmers used to generate deepfakes. The bill was passed by the House of Representatives in December 2019 and is reviewed by the Senate. In June 2021, the House of Representatives proposed a controversial rule that is mandatory for the creator of a deepfake to disclose the media which has been controlled. Each person who produces a deepfake would be required, under the Deepfakes Accountability Act[40], to submit a digital stamp as well as an audio or visual indication of the modification. Any failure to provide the mandatory disclosures or removal of the disclosures would result in a civil penalty of up to $150,000 per incident. Additionally, it would be a criminal violation to intentionally and maliciously omit or delete the required disclosures. Any violation of this regulation could result in a civil penalty of up to $150,000 per incident. The bill would also create a private cause of action, which allows any individual or company whose likeness is used in a deepfake to file a civil suit if the deepfake is not adequate disclosures or if the disclosures are withdrawn. This law will also apply if the deepfake did is not disclosing adequate disclosures or if the disclosures were withdrawn. The non supporters claims that who send dangerous or fraudulent deepfakes are likely to continue doing so anonymously to avoid detection, which is why they are opposed to the measure (and, by extension, liability under the proposed legislation).

The European Parliamentary Research Service's Scientific Foresight Unit (STOA) report on anti-deepfake technology, published in July 2021, addresses the growing concern over the use of deepfakes for malicious

[40] H.R.3230 – 116th Congress (2019-2020): DEEPFAKES Accountability Act of 2021, H.R.3230, 116th Cong. (2021), https://www.congress.gov/bill/167th-congress/house-bill/3230.

purposes.[41] The report presents a comprehensive overview of the current state of deepfake technology and its potential implications, as well as an evaluation of potential solutions for combating deepfakes. The report suggests that developing better detection tools, promoting media literacy, and establishing legal frameworks to regulate the creation and distribution of deepfakes are essential steps to address this issue. Moreover, the report emphasizes the importance of interdisciplinary collaboration and further research to combat deepfakes effectively. This report provides a valuable output for the legislative and research bodies, and anyone concerned about the deepfake technology in society.

False advertising regulations, copyright protections, privacy restrictions, and right of publicity laws, in addition to proceedings for defamation and intentional emotional distress, are the laws that govern deepfakes in the United States of America. Yet, a significant number of these existing legislation include weaknesses that usually prevent victims from obtaining any form of relief. For instance, regulating deepfakes through tort law or copyright infringement law frequently requires the victim portrayed in the image resources, including the time, to bring a suit across jurisdictions and, potentially, against multiple perpetrators, and the victim be able to identify the perpetrator(s) in the first place. Another example would be regulating deepfakes through intellectual property law. In a similar vein, the Communications Decency Act of 1996[42], specifically Section 230, absolves

[41] Mariëtte van Huijstee, Pieter van Boheemen and Djurre Das, "European Parliamentary Research Service", PENAL FOR THE FUTURE OF SCIENCE AND TECHNOLOGY-SCIENTIFIC FORESIGHT UNIT (STOA), July 2021, https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_BRIE(2021)690039(ANN)_EN.pdf

[42] The Communications Decency Act of 1996, No 24, Parliament of Act of 1996, (USA).

producers and users of "interactive computer services" of the bulk of the liability for the information that is given by other information content providers. This indicates that victims may not have a clear way of identifying the perpetrator of a deepfake picture, nor would they be able to take any legal action, for example, a social media or other content sharing site to regulate the usage of such material. Additionally, victims are unable to take legal action against a government agency to prevent the spread of such images. Usage of any United States legislation to control deepfakes will be under scrutiny under the First Amendment. Any law that is currently in place or that will be implemented in the future must be limited to apply only in situations involving true malice or reckless disregard, and where the content is not newsworthy.

Indian law against Deepfakes

According to NCRB Report of 2020[43], the total number of 2756 cases are being registered for offences against women that includes cyber pornography, publishing obscene sexual materials, cyber stalking, cyber-bullying, defamation, indecent representation of women. 1463 cases out of 2756, are cyber pornography resulted cause of revenge.[44] Deepfakes have the potential of committing crimes against persons such as identity theft, blackmail via manipulated video/image (Sextortion), and internet theft, and many others. These are some improper applications of AI-powered technology. Various sections of the Information Technology (Amendment) Act, 2008, in combination with laws of the Indian Penal Code 1860[45] (since the typical

[43] NCRB (National Crime Records Bureau) Report 2020.

[44] NCRB (National Crime Records Bureau) Report 2020.

[45] Indian Penal Code of 1860, No. 45, Acts of Parliament of 1860 (India).

purpose is to conduct identity theft, blackmail through manipulated video/image, or online theft) may be used to address this issue. In India, there is no law pertaining to Deepfake-related offenses. Government efforts to combat this burgeoning crime are insufficient. According to the Indian Penal Code 1860, the accused might be charged with defamation (Sections 499 and 500). When a Deepfake video or audio of a person is made in which he seems to say anything damaging to that person's reputation, the criminal defamation statute can be applied. This category may include a bogus video in which a person says something unsettling. As established in Sunilakhya v. HM Jadwet[46], the intention to do injury to a person's reputation is a necessary element of a defamation offense under criminal law. Deepfakes constitute defamation, which includes any visual portrayal.

The concept of obscenity is usually considered as violation of community standard and public decency and which is repulsive and prurient to society. Within the purview of Indian legal regime, section 292 of India Penal Code [47]specifically provide punishment for sale, distribution, importation/exportation of obscene material. In the same way, there is punishment provision for dissemination of obscene content in electronic form as mentioned in the section 67 of Information Technology Act, 2000.[48]

Section 66E of the Information Technology Act 2000 establishes penalties for privacy violations. Similarly, the sections 67Aand 67B of Information Technology Act, 2000 penalize the transmission or publication of sexually explicit content, or sexually explicit depictions of children by electronic means, respectively. Identity theft entails the use of fraudulent or deceptive

[46] Sunilakhya v. HM Jadwet, AIR 1968 Cal 266.

[47] The Indian Penal Code,1860, § 292, No. 45, Acts of Parliament of 1860 (India).

[48] Information Technology Act, 2000, § 67, No. 21, Acts of Parliament of 2000, (India).

means to steal an individual's identity details for gaining access to resources or obtaining credit and other advantages in the victim's name. Even, the IT Act, 2000 also consider cheating through personation by using computer facility as penal provision mentioned in its section 66D. It is important that social media intermediaries must implement the establishment of self-regulating body to address any grievances(if any) and supervise the follow-up of code of ethics as mentioned in the section 11 of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021.[49] It is also expected from social media intermediary to formulate a due diligence document and regulation where explicit information must be shared with users of computer resource regarding prohibition of uploading or sharing any obscene, prurient or pornographic content. An intermediary may terminate the access of concerned user in case of non-compliance of such regulations by any user.[50] Recently, a proposed Digital India Act, 2023 draft was discussed by Ministry of Electronics and Information Technology wherein there is discussion of taking appropriate action against users who are involved in revenge porn, cyber-bullying. There is reference of conventional quality testing mechanism of risk prone AI based technology in the interest of supervision of digital content and content moderation on periodic basis.[51]

[49] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, § 11, No. 21, Acts of Parliament of 2021, (India).

[50] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, § 3, No. 21, Acts of Parliament of 2021, (India).

[51] Proposed Digital India Act 2023, Digital Indira Dialogues, 9th March,2023, MINISTRTY OF ELECTRONICS AND INFORMATION TECHNOLOGY, GOVERNMENT OF INDIA, https://www.meity.gov.in/writereaddata/files/DIA_Presentation%2009.03.2023%20Final.pdf

Recently Supreme Court of India Suo Moto[52] under the quorum of Hon'ble Mr. Justice Madan B. Lokur and Hon'ble Mr. Justice Uday Umesh Lalit, an application was filed with a brief affidavit from the Ministry of Home Affairs to eradicate child pornography, rape, and gang rape images, videos, and sites in content hosting platforms and other apps, the legislative body is responsible to draft and adopt guidelines. The order was passed through which law enforcement bodies can collect evidence to prohibit the websites including child pornography through deepfakes or various means. It is challenging for the administration and court to address such concerns and new crimes[53]. Such crimes are convenient[54] and transnational. They are swift and hard to track down. Because of established conventions, poorly governed severe rules, and imprecise, confusing definitions, they are immensely lucrative. On a worldwide scale, there is a growing awareness of this technology, and corporations and research institutions have taken things by doing research on strategies to delete such information online. Because of this, there has been considerable success, although technical crime is sometimes two steps ahead of the measures developed to counteract it. In view of this expanding threat, forensic analytical methodologies, standards of evidence, and relevant legislative mechanisms would need to be reexamined.

[52] In Sexual Violence Videos and Recommendations, (CRL) No (s). 3/2015 PRAJWALA Letter dated 18.2.2015.

[53] Hoar, Identity Theft: The Crime of the New Millennium, 80 Oregon L. R., 1421 (2001)

[54]McCusker, 'Transnational Organized Cybercrime: Distinguishing Threat from Reality'" CRIME, LAW AND SOCIAL CHANGE 2006.

Conclusion and Suggestions

With the rise of technological advancement and innovation, there is always existence of adverse repercussions in the form of digital pornography, indecent representation through morphing videos, pictures. It realizes that deepfakes as technology is double-edged sword. Any type of vitriolic distortion and manipulation of video and images are possible through use of deepfake technology to cause ignominy for victim which may be either in the form slut-shaming or revenge porn etc. It is true fact that cyber criminals find it easy to adapt these AI based technologies leading to major technological battle between cyber criminals and law enforcement agencies. Sometimes, it is difficult to identify, track or gather evidences against cyber criminals who are involved in the transmission of obscene and pornographic content using deepfakes technology in the internet resources mainly because they work using false IP address anonymously or using pseudo-name. The Government can strategize and formulate a model code of conduct against fake news, misinformation and obscenity for social media intermediaries and also put responsibility on deepfakes creating companies.

The malicious approach of cyber criminals violates not only the right of privacy of individuals, but also result into disparaging acts. The existing Indian legal regime is inadequate to discover issues related with AI algorithms. Due to lack effective regulations, there is rise in the cases of misuse of AI based technology as it is convenient for cyber criminals to replace facial expression or morph the existing videos with malicious intent. It can be created with intent to spread false news or use for financial misappropriation or depiction of pornographic content or hate speech. It is important for the government to adopt strict censorship measures whereby specific order may be given to intermediaries for completely blocking access

of any obscene, pornographic content and impose penalty wherever it is necessary for preventive and deterrent effect. The Government can also effectively use the provision of section 69A of the IT Act,2000 by giving directions for blockage of any information for public accessibility in the interest of public order.

There is no doubt about the fact that India's legal framework for technology is insufficient to adequately address the challenges posed by artificial intelligence algorithms. In this regard, the following recommendations are necessary to tackle the misuse of deepfake technology in the rise of cybercrime against women: Implement anti-fake technologies. - Governments may protect themselves from deepfake assaults by developing a dependable method for identifying them, particularly by leveraging automated solutions and help in battling this war against deepfakes for women and children protection. An example of possibility is AI-powered detection software. Using the same deep learning techniques, one may also create deepfakes to identify indications of image or video manipulation. The detecting deepfakes, and this is only one of them. Watermarking information to identify tampering is an additional technological method. For instance, Amber Authenticate is a cryptographic gadget capable of generating hashes at predefined times during a movie. If the movie is modified, the hashes will change, signaling to the user that the content has been altered. Training and awareness- The threat's uniqueness is among the reasons why deepfakes pose a risk to individuals. Mostly people are still unaware of deepfakes and their destructive potential. Government can reduce their risk of falling prey to a deepfake attack by notifying their departments about the threat. Apart from effective implementation of existing relevant laws, it is important to expand the horizon of public awareness against the misuse of such type of technology or its similar kind. Such level of sensitization needs a conscious collaboration

of media, civic society, government to understand and realize their responsibilities. It's possible that technology is not the sole approach to avoid deepfake videos. Surprisingly effective are fundamental security best practices for combating deepfake. Numerous deepfakes and associated scams would have been avoided if automated checks had been included into money disbursement processes, for example, educate others on how to spot a deepfake. Ensure that media literate and utilize credible news sources. Establish superior foundational procedures — "trust but check."

People should explore the usage of blockchain technology. The use of blockchain technology as a solution might be a realistic approach. A decentralized system, often known as a blockchain, enables users to save data online without using centralized servers. Moreover, blockchains are immune to provide wide variety of security flaws sensitive to centralized data storage. Distributed ledgers are appropriate for storing hashes and digital signatures, notwithstanding their inability to store vast volumes of data. Individuals may, for example, utilize blockchain to authenticate and verify the legitimacy of a personal video or audio file. The electronic signatures on a film, the more likely it is to be considered a genuine record. This is not the optimal choice to To analyze and account for the competency of individuals who vote on a file, extra procedures will be required. We should look doe adopting a zero-trust stance towards internet material- Using deepfake software, con artists may go one step farther than simply stealing photographs and creating fake accounts on the Internet. Before putting your confidence in anything you see on the Internet, take a zero-trust stance and exercise extreme vigilance. This recommendation applies to all fresh narratives, photos, and videos. Ensure that individuals are prepared to respond appropriately to a deepfake. Prepare a plan of action that can be implemented if a deepfake is detected. Identity checks, such as validation, device ID and analytics, behavioral analytics, and