The Indian Journal for Research in Law and Management

Open Access Law Journal – Copyright © 2024

Editor-in-Chief – Prof. (Dr.) Muktai Deb Chavan; Publisher – Alden Vas; ISSN: 2583-9896

THE ERA WHEREIN EYES CANNOT BE DEPENDABLE: IMAGE MANIPULATION BY MEANS OF DEEPFAKES

ABSTRACT:

Deepfake, a deep learning-based method, facilitates image and video manipulation. Visual evidence is often used in investigations and legal proceedings. These particular pieces of documentation may now be suspicious as a result of technology advances, notably Deepfake. Edited photographs and videos are extremely lifelike and impossible to distinguish from the originals. Deepfakes have been used to coerce people in general, plan terrorist attacks, spread false information, discredit individuals, and cause political unrest.

In recent years, digital revolution has normalised internet-based reality distortion. It took a quantum leap with the introduction of 'deepfake' technology. Machine learning now allows for the alteration of material to create hyper-realistic fake content that is resistant to detection. Deepfakes offer advantages, but they can pose hazards that outweigh these benefits.

This study seeks to explore the present legal environment for deepfakes. We intend to present an examination of whether India's current legal system is capable of dealing with the unique issue of deepfakes. A wide range of legal rules are examined in relation to the supervision of deepfakes. This research tries to uncover the accessible solutions available for the misuse of deepfakes.

This article aims to provide readers with a deeper understanding of deepfakes, including how they are formed and detected, recent advances, weaknesses in existing security measures, and topics for further exploration and attention.

Keywords: Deepfakes, fake news, technological remedies, artificial intelligence.

1. INTRODUCTION

Deepfakes have become practically indistinguishable from actual footage over the last year. AI and machine learning will drive further advancements in technology. As greater numbers of people discover how to use deepfake technology, Videos will grow increasingly popular and begin to enter the public psyche. As fakes become more prevalent, it becomes harder to discern between legitimate and doctored films, potentially leading to legal, social, and political damage in several sectors of our everyday lives. Deepfakes remain uncontrolled in 2019, with no clear legal framework in place.

Deepfake technology is the most recent addition to the collection of image modification technologies. Deepfake is a software that uses machine learning techniques to add voices and faces to audio and video recordings that effectively pull off lifelike impersonations.[1]

Filmmakers' science fiction helped popularize artificial intelligence in the early 20th century.[2] Nowadays, technology is used in several fields, including medical and social media with Artificial Intelligence. Machine learning, a subtype of artificial intelligence, allows humans to use technology to resemble human intellect, revolutionizing the way it functions. For decades, audio-visual manipulation has dominated the film business. Machine learning and automation have eliminated the tedious and time-consuming tasks associated with technology manipulation. Technological breakthroughs have turned what was once considered fun into a threat to society.

Thus, Deepfake manipulates or synthesizes voices, faces, and even emotions to produce fake material. It mimics a genuine person, yet it does things that the real person did not do. The world is not new to altered visuals. When photography first emerged in the first half of the 1800s, the most popular methods for manipulating photographs were manual methods or light effects. Images were edited using paint, which is ink, or similar methods before the advent of Photoshop. Photoshop made it possible to significantly alter photographs.

1.1 What Is a Deepfake?

[1] Cole S., 'Deepfakes were Created as a Way to Own Women's Bodies—We Can't Forget That,' (VICE, June 18,2020) https://www.vice.com/en/article/nekqmd/deepfake-porn-origins-sex-ism-reddit-v25n2

[2] Anyoha R., 'The History of Artificial Intelligence' (Science in the News, April 23, 2020) The History of Artificial Intelligence - Science in the News (harvard.edu)

A deepfake is a very realistic digital fabrication of photos, videos, and sounds, combining the terms "deep learning" and "fake."[3] Simply said, a deepfake is a manufactured film that represents an event that never occurred by modifying previously existent video footage or images.[4]

Deepfakes may range from stupid to menacing. Deepfakes have been used for a variety of purposes, including inserting Nicolas Cage in famous movie sequences like Raiders of the Ark of the Lost and showing a Wall Street Journal writer imitating Bruno Mars' dance steps.[5] However, because deepfakes' beginnings are intimately related to sexually explicit material, a darker point of Many deepfakes focus on making pornographic movies of well-known celebrities. A frightening example of deepfakes was a video showing assault weapons activist Emma Gonzalez ripping up an original of the Constitution[6]. The original video showed Gonzalez advocating for gun control by breaking up a target. However, the image was modified for offensive motives.

2. THE HISTORIC AND TECHNOLOGICAL BACKGROUND OF DEEPFAKES

Humans have been experimenting with photography for almost as long as it has existed[7]. One early example is a famous portrait of Abraham Lincoln from 1860. Although the image seems legitimate, it is actually a composite of pictures of Abraham Lincoln's head and John Calhoun's torso[8]. However, the popular picture editing software Photoshop is the most widely recognized example of picture alteration technology. Photoshop was created in 1987

[3] “Deep learning” refers to a branch of artificial intelligence where software learns how to recognize patterns out of data. The software learns “in a very real sense” by mimicking how the brain utilizes neurons to think. Robert D. Hof, Deep Learning, MIT TECH. REV. (Apr. 23, 2013), https://www.technologyreview.com/s/513696/deep-learning

[4] John Brandon, Terrifying High-Tech Porn: Creepy ‘Deepfake’ Videos Are on the Rise, FOX NEWS (Feb. 16, 2018), http:// www.foxnews.com/tech/2018/02/16/terrifying-high-tech-porn-creepy-deepfake-videos-are-on-rise.html ; see also Bobby Chesney & Danielle Citron, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, 107 CALIF. L. REV. (forthcoming 2019) (manuscript at 4) (on file with author).

[5] Hilke Schellmann, Deepfake Videos Are Getting Real and That's a Problem, WALL ST. J. (Oct. 15, 2018, 5:29 AM), https:// www.wsj.com/articles/deepfake-videos-are-ruining-lives-is-democracy-next-1539595787

[6] See Chesney & Citron, Gianluca Mezzofiore, No, Emma Gonzalez Did Not Tear Up a Photo of the Constitution, CNN (Mar. 26, 2018, 3:30 PM), https://www.cnn.com/2018/03/26/us/emma-gonzalez-photo-doctored-trnd/index.html.

[7] See Megan Garber, Oprah's Head, Ann-Margaret's Body: A Brief History of Pre-Photoshop Fakery, ATLANTIC (June 11, 2012), https://www.theatlantic.com/technology/archive/2012/06/oprahs-head-ann-margarets-body-a-brief-history-of-prephotoshop-fakery/258369/.

[8] Photo Tampering Throughout History, GA. TECH. C. COMPUTING, https://www.cc.gatech.edu/~beki/cs4001/history.pdf

and became widely available by 1990. Photoshop is a popular program among photographers for manipulating images, including magazine covers and Instagram postings.[9]

Deepfake replaces the face of the "target" in the input video with someone else's. It is made via neural networks, a type of machine learning that is trained to "map the facial movements of the source" when making the synthetic movie. Generative Adversarial Networks, or GAN, is a novel technique that combines two neural networks.

The proliferation of deepfake videos highlights the technology's fast growth and its downsides. Elf Yourself, an Internet craze where individuals put images of themselves into a prepared movie of Christmas elves performing to Christmas tunes, might be regarded as the divine ancestor of deepfakes. The ElfYourself films are clearly phony, as seen by the superimposition of elf heads on artificial bodies, despite similarities to deepfakes[10].

The most recent version of hyper-realistic, generated deepfakes originated on the social media platform Reddit. The first authentic deepfake was discovered on the r/CelebFakes subreddit, which specializes in photoshopping celebrities to make them seem naked[11].

3. THREATS OF DEEPFAKES

Rapid innovation has sped up organizational change, but it frequently comes at the cost of security. Businesses are developing, but the margin for mistake is increasing. This is digital Acceleration has widened the attack zone and increased the amount of assets that require enhanced protection, placing enterprises at danger. Currently, authentication capabilities are insufficient to address the growing threat of deepfake technology.

While innovative leadership is important, security must also be considered during the transition process. Cyber mishaps are becoming increasingly often due to advancements in technology and increased sophistication. A more data-driven political danger landscape. Rapid innovation might leave security behind in company transformation efforts. 39% of worldwide IT leaders say their organization's security procedures have not kept up with digital transformation projects.[12]

[9] See Garber, supra note 8.

[10] ELFYOURSELF, https://www.elfyourself.com/

[11] Aja Romano, Why Reddit's Face-Swapping Celebrity Porn Craze is a Harbinger of Dystopia, VOX (Feb. 7, 2018, 5:55 PM), https:// www.vox.com/2018/1/31/16932264/reddit-celebrity-porn-face-swapping-dystopia

[12] IT Leader Survey, Veritas Technologies, 2021.

Deepfakes proliferate online due to several circumstances, including information cascades. Information cascades arise when individuals share information and assume its credibility based on the number of shares. People's predisposition to believe what others say, even if it contradicts their own opinions or expertise, contributes to the information cascade. Similar to a beehive, social media connects everyone and everything. Complex relationships increase the likelihood of direct influence between individuals. Mindless imitation is not necessarily illogical. It is often based on reasonable interpretation of inadequate knowledge, which might cause more harm than a lack of information.

3.1.Non-Consensual Pornography

Deepfakes have the potential to do major harm in a number of different areas, such as revenge pornography, disinformation, and the reliance on video as a medium, as the technology underlying them develops quickly. As will be covered in more detail in the Comment, deepfakes may have more negative effects than positive ones.

Since deepfakes have been used to produce celebrity porn, it is conceivable that dishonest performers would also utilize the technology to produce revenge porn for those who are not well-known. Revenge porn is defined as "The distribution of explicit sexually explicit pictures or recordings of somebody else with that person's consent or knowledge."[13] It is sometimes referred to as "involuntary porn" or "nonconsensual pornography."[14] Distribution of graphic images or films obtained without permission, consented to but with an awareness of privacy, or produced through "sexualized photoshopping" are all examples of revenge porn.[15] The emergence of deepfakes has made it possible to "sexualize photoshopping" for both photographs and films. Due in part to the popularity of revenge porn, "the majority of victims of fake sex videos will be female."

In the guise of a deepfake, vengeance porn has already affected at least one prominent person. Before anonymous predators altered her face onto explicit photographs of another person's body, Noelle Martin of Perth, Australia, had already suffered years of being the target of revenge porn.

[13] Aja Romano, Why Reddit's Face-Swapping Celebrity Porn Craze is a Harbinger of Dystopia, VOX (Feb. 7, 2018, 5:55 PM), https:// www.vox.com/2018/1/31/16932264/reddit-celebrity-porn-face-swapping-dystopia

[14] Caroline Drinnon, When Fame Takes Away the Right to Privacy in One's Body: Revenge Porn and Tort Remedies for Public Figures, 25 WM. & MARY J. WOMEN & L. 209, 211 (2017).

[15] Amanda L. Cecil, Taking Back the Internet: Imposing Civil Liability on Interactive Computer Services in an Attempt to Provide an Adequate Remedy to Victims of Nonconsensual Pornography, 71 WASH. & LEE L. REV. 2513, 2520 (2014).

"Pornography" refers to the representation of sexual behaviors in literature, films, or other media to elicit sexual enjoyment. Pornographic websites, computer-generated content, and internet surfing can all lead to the transmission of pornographic material. This category includes films, writings, images, and other media. Individually viewing porn in India is not considered an offense under Indian Penal Code. However, the Indian judiciary has established constraints on the freedom to watch pornography.

A crime such as revenge pornography is no different from rape, and it is extremely hard for the person being targeted to resist when the offender betrays their trust and uses blackmail to further their own agenda, ruining the family's name in the community. The court in State of W.B. v Animesh Box[16] saw the victim of revenge porn as a survivor of rape and awarded the victim suitable compensation. In this day of rapid technological development, even our personal electronics are no longer safe.

3.2.The component of sexual privacy

Sexual privacy is a component of a person's right to privacy, which safeguards an individual's dignity and preserves their sense of self. It guarantees personal autonomy, which identifies a person.

Since it determines sexual agency, closeness, and equality, sexual privacy is the highest value in privacy. The personal decisions made in intimate relationships, such as if or not to reveal one's body to another, are covered by sexual privacy. A person may control and establish borders around their body with the help of sexual privacy. It governs how one experiments with gender identity, sexuality, and sexual choices.[17]

In Griswold v. State of Connecticut[18], the court recognized one's right to sexual privacy and emphasized the significance of privacy in relation to birth control tablets.

In the case of Francis Coralie v. UT of Delhi, the Supreme Court ruled that the right to human dignity encompasses the basic requirements of life, such as sufficient food, clothes, and other essentials.[19] As a result, court interpretations have expanded the definition of a dignified existence that encompasses a wide range of rights.

[16] R.M. No. 11806 of 2017, GR/1587/2017.

[17] Citron D., 'Sexual Privacy' (2019) 127 (7) https://www.yalelawjournal/ The Yale Law Journal

[18] 1965 SCC OnLine US SC 124: 14 L Ed 2d 510: 381 US 479 (1965)

[19] (1981) 1 SCC 608: AIR 1981 SC 746: (1981)2 SCR 516.

The right to privacy is a constitutionally protected right that arises in various contexts that are ensured in Part III of the Constitution of India, but primarily from the right to life and personal liberty enshrined in Article 21, according to Justice D Y Chandrachud's historic ruling in K.S. Puttaswamy v Union of India.[20] "The fundamental basis of human dignity is privacy. There are normative and descriptive purposes for privacy. Privacy supports the timeless values that form the foundation of the guarantees of life, liberty, and independence on a normative level. Descriptively speaking, privacy assumes a range of rights and interests that form the cornerstone of lawful liberty. The court set out on a lengthy road to acknowledge the diversity and plurality of privacy concepts.

Deepfakes might potentially be effectively addressed by the law in the field of copyright infringement.[21] However, going that approach would need that the victim of the deepfake originally took the video. Furthermore, it's possible that the owner of the original video is not the same individual that the deepfake injured in reality. Moreover, the author of a deepfake could be able to argue that deepfakes are fair use in court, even if the individual who was affected owns a copyright.[22] The transformative aim of copyrighted content is a crucial distinction in determining whether or not its utilization by others qualifies as fair use, even if a detailed discussion of infringement of copyright and fair use is outside the purview of this Comment.

4. PRESENT LEGAL SITUATION

The Deepfakes Accountability (Bill) Act 2019, which sadly failed to pass the Parliament, was an attempt by the United States to combat deepfakes. The law suggested protective measures for those harmed as well as penalties for failing to watermark deeply fabricated information to indicate that it is changed or fraudulent.

Indian laws do not currently have sufficient mechanisms in place to identify or control deepfakes. As far as deepfakes are concerned, the Information Technology Act of 2000 is the best available remedy under Indian law. In the lack of particular regulation, the IT Act is the

[20] (2017) 10 SCC 1.

[21] Megan Farokhmanesh, Is It Legal to Swap Someone's Face into Porn Without Consent? VERGE (Jan. 30, 2018, 2:39 PM), https:// www.theverge.com/2018/1/30/16945494/deepfakes-porn-face-swap-legal.

[22] See Roberts, David Greene, We Don't Need New Laws for Faked Videos, We Already Have Them, ELECTRONIC FRONTIER FOUND. (Feb. 13, 2018), https://www.eff.org/deeplinks/2018/02/we-dont-need-new-laws-faked-videos-we-alreadyhave-them

strongest line of defense against deepfake cybercrimes, however its efficacy in combating them is debatable.

Under section 52 of the Indian Copyright Act[23], which permits the author restricted use of copyright material without the consent of the first author, a deepfake may be justified or defended in the Indian context. The notion of fair use of copyright work in the United States mostly pertains to labor-intensive and skill-based transformative activity. In contrast to the fair use notion, Indian copyright law leans more toward the idea of "fair dealing," which is specified in Section 52 of the Copyright Act, 1957 but is not specifically defined. This section extensively describes several types of fair dealing.

Because deepfakes are not included in the specified exceptions, it is easier to hold the developer accountable. Furthermore, Section 57 (1) (b) provides protection against distortion, mutilation, and alteration of copyrighted material.[24] The inflexibility of the equitable Deepfakes may be effectively addressed with the use of the dealing idea.

5. THE SOLUTIONS FOR DEEPFAKES

"Technologies that can be exploited to distort and augment reality are developing more quickly than our comprehension, control, or mitigation capabilities."[25] With deepfakes' technology becoming more sophisticated, it could be already too late for any limitations to have a significant impact. Regulating the movies that are really made is still possible, even if it is probably too late to govern the technology underlying deepfakes. Additionally, it is during this time unsure of how to respond to deepfakes that have previously been reported yet violate the new rules. Since America lacks rules pertaining to the "right to be forgotten" for material uploaded online, erasing an existing deepfake may provide further challenges.[26]

Deepfakes are governed by the General Data Protection Regulation (GDPR) of the European Union.[27] The GDPR may be used to deepfakes as it offers protection and remedies for processing data. Include data manipulation. According to the GDPR, a data subject has the

[23] Copyright Act, 1957, s. 52.

[24] Copyright Act, 1957, s. 57 (1)

[25] Charlie Werzel, He Predicted the 2016 Fake News Crisis. Now He's Worried about an Information Apocalypse, BUZZFEED NEWS (Feb. 11, 2018, 8:45 PM), https://www.buzzfeednews.com/article/charliewarzel/the-terrifying-future-of-fake-news.

[26] See Emma Grey Ellis, People Can Put Your Face on Porn--And the Law Can't Help You, WIRED, https://www.wired.com/story/face-swap-porn-legal-limbo/.

[27] General Data Protection Regulation (2016) OJ L 199/1.

right to have their personal information removed from internet searches that they don't anymore want to be shown. This is known by the ability to be forgotten.

According to Article 5 of the EU General Data Protection Regulation, deepfake material needs to be removed right away if it is false, irrelevant, or erroneous. Furthermore, the "right to erasure"—which is afforded to residents of Europe under Article 17—may be used by a data subject who has been the victim of a deepfake, regardless of whether the information is factual or correct.

As a component of the right to privacy under Article 21, the right to be forgotten has been acknowledged as a basic right in the Indian context. Beyond the protection of privacy in terms of location and physical aspects, the right to privacy is now extended to include personal data. The seminal ruling in K.S. Puttaswamy v Union of India acknowledged the extension of the idea of privacy to encompass informational privacy. Following this, the Karnataka High Court noted in Vasunathan v. High Court[28] that "the being modest and reputation of the people that are involved, especially if it encompasses women, cannot be made available to everyone indiscriminately," even though the court held that the right to be forgotten must be implemented in sensitive cases.

A fair and impartial regulatory reaction would have to take into account the interests of the parties involved as well as the benefits and drawbacks that deepfake brings about. The main parties involved in deepfake. The victim is the main application, but there are other maker, spectator, and disseminator roles. The legislation should permit the use of deepfakes for legitimate reasons but outlawing them for other uses, as opposed to outright forbidding them. When granting the creator of a deepfake control over their material, this criterion for the victims must be weighed against their own. The most crucial tenet would be to guarantee that these prerogatives are protected by appropriate remedies.

6. CONCLUSION

As we've seen, the reliability of video evidence has only recently been compromised by technical advancements, despite the lengthy history of picture and even video tampering.

[28] 2017 SCC Online Kar 424.

Contemporary Deepfake software may quickly produce films that appear absolutely authentic to the untrained eye. Furthermore, if attackers can test and alter the faked video as needed, automated systems may be readily taken down.

Given a result, education and media literacy continue to be the most effective means of combating misinformation, given there is no apparent simple technological answer and misinformation generally can be attributed to these factors.

The boundary between the genuine and the false is becoming more hazy because to the quick rise in the breadth, scale, and complexity of digitization. On the internet, digital disinformation presented as altered images or fake news has taken on a life of its own. Deepfakes' legal murkiness might make them appear like a threat to civilization. It is important to remember that deepfake technologies were first developed to imagine and re-create images of how particular celebrities would appear in awkward situations. Later, tech companies used the technology to create learning tools for therapeutic and educational purposes, and some historians even used it to recreate historical figures for use in video games and other artistic mediums.

People will gradually become more conscious of the fact that they shouldn't always accept what they see as deepfake videos become more common.

Deepfakes may now go in two directions: either they can be beneficial or harmful, just like any other machine. In order to fully profit from deepfakes, we must initially take regulatory action to lessen their risks.