THE ILLUSION OF AUTHENTICITY: MAPPING DEEPFAKE TECHNOLOGY IN THE DIGITAL AGE

Pooja Yadav, B.A. LL.B. (Hons.), Amity Law School, Noida,
Amity University, Uttar Pradesh

ABSTRACT

The Artificial Intelligence is one of the rapid technological advancements, which has transformed the digital innovation into a powerful tool that could generate highly-realistic synthetic media, commonly termed as "deepfakes". The emergence of AI-based deepfakes resulted in a paradigm shift in public perception, which means while people were thinking about the positive role of technological advancement in providing aid and innovation, the issue of technology is viewed now as a mechanism of deception, misleading, manipulation, and rights violation, raising various legal concerns. This paper showcase how evolution of Artificial technology resulting deepfakes, has gone from innovation to manipulation. This further represents how the nature of deepfakes, the models especially Generative Adversarial Networks (GANs), used to make the deepfakes that appears to be hyper-realistic replica of an individual. Additionally, the benefits of deepfake technology in several domains, and the misuse of such deepfake technology that has overshadowed its benefits, as misuse is raising significant questions as to the privacy and personal dignity of an individual. Hence, questioning the adequacy of current laws in addressing the issues and challenges arising from the deepfake technology.

Keywords: Artificial, intelligence, deepfake, technology, innovation, laws, manipulation, public rights, etc.

1.1 INTRODUCTION

The nature of human beings is rooted in the relentless curiosity that drives them to learn, explore, innovate, and more. As pointed out by Aristotle, "All men by nature desire to know", as curiosity is a fundamental, universal human trait, internally visualizing something that doesn't even exist, and then ultimately creating it. Thus, looking back to the earliest stages of human progress, the tenacious yearning to create machines, systems, advancements that can ease the doing of the human task, enhance efficiency, simplify the complexities faced in a regular life, etc. In the late 1600s, a German attorney, G.W. Leibniz, theorized that machines would someday use a binary system to calculate numbers, as according to him, it is unworthy of excellent men to lose hours like slaves in the labour of calculation, which could safely be relegated to anyone else if machines were used. With such an aspiration to ease of doing and saving time and effort, there have been rapid technological advancements, and one of these is Artificial Intelligence, which was developed with the main aim to transform industries, improve decision-making, enhance efficiency, saving time and effort in almost every diverse profession. However, due to its easy accessibility and usage, its application reaches far beyond mere assistance, to the creation of replication of human face, voice, manner of speech, and other personal attributes, thereby evolving deepfake technology in between such transformation and advancements.

This newly evolved Deepfake Technology covers a broad spectrum of techniques such as creation, alteration, and manipulation of data and media through automated processes. The term "deepfake" in itself is a portmanteau of deep learning and fake. Basically, deepfakes are the AI generated videos, pictures, or audio clips that look real. They can be used for fun, or even for scientific research, but sometimes they're used to impersonate people like politicians or world leaders, in order to deliberately mislead people[1].

1.2 EVOLUTION OF DEEPFAKES

Deepfake AI is a new form of technology having its origins on the manipulation of photos and videos through several programs. Regardless, of being the new form of technology, it can still be traced back to the 1900s wherein, the researcher, and scholars at Academic Institutions begin to explore the use of AI for image processing. During Mid 2010s, the cheap computing powers,

[1] BBC Newsround, Deepfake Technology: What Is It, How Does It Work, and What Can It Be Used For?, BBC (May 15, 2024)

big data sets, AI, and machine learning technology were combined to make and improve the cultured deep leaning algorithms.

In 2014, University of Montreal Researcher Ian Goodfellow, developed the GAN, which is the heart of deepfakes[2]. However, the term 'deepfake' was still not existed till 2017, where anonymous Reddit user named "deepfakes" began realizing deepfake videos of the celebrities and the face swapped images or videos as the face of a person of swapped with another, by using the GAN tool, going viral on the internet and social media platforms.

These chain of incidents making the deepfake content popular, led the big tech companies like Facebook, Google, and Microsoft to invest in the development of tool which can be used to detect the deepfakes[3]. However, regardless of numerous attempts by the government, private companies to combat and detect deepfakes, but the technology further continues to advance and produce more AI generated deepfake images and videos which are convincingly look like real[4].

1.3 CONCEPT AND NATURE OF DEEPFAKES

Deepfakes are a subclass of synthetic media that are AI-driven images, videos, or audio clips that show something that does not exist in reality or events that have never occurred.[5] The term came to be used for synthetic media in 2017 when a Reddit moderator created a subreddit called "r/deepfakes" and began posting videos that used face-swapping technology to insert celebrities' likenesses into existing pornographic videos[6].

How deepfakes are made- Deepfakes are not the edited videos or pictures, but they are the newly created one by using the specialised algorithms, for instance, the refined facial features are analyzed through Machine Learning in order to manipulate them within the context of other videos.

Generative Adversarial Networks (GANs) are often used to produce deepfakes. It comprises of two different yet complementary AI deep-learning model which work together for the

[2] Ian J. Goodfellow and others, 'Generative Adversarial Nets' in *Advances in Neural Information Processing Systems 27* (2014) 2672.
[3] Cade Metz, Tech Giants Race to Detect 'Deepfake' Videos' *New York Times* (2 January 2020).
[4] Bobby Chesney & Danielle Citron, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107 *California Law Review* 1753, 1756–57.
[5] Laura Payne, Deepfake, *Encyclopaedia Britannica* (31 March 2026).
[6] TechTarget, What Is Deepfake Technology? TechTarget

creation of highly realistic yet fabricated material which appears to be real but actually stands false, i.e., deepfake. Out of these two models, one is used to create a replica of real image or video, and the other detects flaws in such replica, ensuring if it is fake, thereby detecting the errors or any differences between the original and replica, and if any exists, the same is being reported. The same process is continued multiple times until the other model detect no flaw or difference between original and replica[7]. Thus, in other words, the first model produces the deepfakes and receives feedback from the other model, helping to adjust and make it look more real. The same process is repeatedly followed until the other model does not detect any falsehood in the creation. These two models can also be regarded as two algorithms namely – a generator and a discriminator, i.e., the one which creates the content (deepfake), and the other one which detect or refine the fake content.

The GAN system scrutinizes the videos or the pictures from different angles, deeply analysing the behaviour, movement, body language, and the speech patterns to make the deepfakes more realistic, leaving no signs of being fake. Such deepfakes can be created on one of the two ways, where one can use the original videos or images of a concerned person, wherein such person is made to say or do things which he never said or did, and second, wherein, one swaps the face of the person onto a video or images of another person, this is known as face swap[8].

There are certain specific approaches that are used to create deepfakes:

- **Source Video deepfakes-**

  The original, real images or videos are used as the base for the manipulation. A neutral network based deepfake autoencoder is used to analyse and understand the relevant traits of the targeted person, such as facial expressions, body language, manner of speech, etc, and then these characteristics are imposed onto the original videos or images. These attributes are encoded by the autoencoder and decoder, then imposed onto the targeted videos.

  One of the perfect examples of this deepfake is the Rashmika Mandana case[9], wherein

[7] Machine Learning- Ian J Goodfellow and others, 'Generative Adversarial Nets' in *Advances in Neural Information Processing Systems 27* (2014) 2672
[8] BBC Newsround, What Are Deepfakes and How Do They Work? BBC (May 15, 2024)
[9] Soumya Gupta, 'Rashmika Mandanna Deepfake Video Sparks Debate on AI Misuse in India' *Indian Express* (November 2023)

the original video and image of her was used as the base, and her face was digitally superimposed using AI, creating a look-alike, realistic but fabricated video of her spread on social media, while retaining the original body, movements, and facial expressions.

- **Audio Deepfakes-**

  This technique is generally used by video game developers. This is developed by GAN, where it clones the audio of a person's voice, and on the basis of these vocal patterns, an AI-generated model is created to make him say whatever the creator wishes to say.[10] These are commonly used in blackmail, scams, and fraud, threatening someone to pay money, extracting personal information, manipulating decisions, etc.

  These are often used in political campaigns as well, for instance, during the 2024 US presidential election, robocalls using the cloned voice of Joe Biden misled people by delivering fabricated messages[11].

- **Lip Syncing-**

  This is another technique used in deepfakes, supported by recurrent neural networks.[12] This can be regarded as the next step to audio deepfake, as the person's lip movements are altered in a manner to match the fabricated audio. Consequently, making the person appear to say something which he never said. One of the notable examples is the circulated videos of Volodymyr Zelenskyy, wherein he was seen announcing surrender during the Russia-Ukraine War[13]. Such deepfakes wherein lip-syncing is used to fabricate the statements which appears that the concerned person has actually made those statements, whereas on the contrary such person in reality has never said the same. Thereby, disseminating manipulative and fabricated content to mislead and influence the public perception.

[10] Ian J Goodfellow and others, 'Generative Adversarial Nets' in *Advances in Neural Information Processing Systems 27* (2014) 2672.
[11] Tiffany Hsu and Steven Lee Myers, 'Fake Biden Robocall Targets Voters Ahead of Primary' *New York Times* (January 2024).
[12] Laura Payne, 'Deepfake' *Encyclopaedia Britannica* (31 March 2026)
[13] Kevin Roose, 'A Fake Video of Zelenskyy Appears to Show Him Urging Ukrainians to Surrender' *New York Times* (16 March 2022).

1.4 INNOVATION VS MISUSE

The advent of deepfakes represents the rapid progress in the AI- generated content, striking the perfect example as to how innovation can, at the same time, enhance and endanger society. Rooted in AI and deep learning, deepfakes, especially through neural networks, create the super-realistic synthetic media, such as videos, images, and audio that are the replication of the human appearance, and speech, body language, and other consisting other attributes. These often pose solemn threat to the privacy, democracy, and legal system, etc. consequently, this dual character makes deepfakes as one of the most debated technological developments in the contemporary digital world, where people are praising the same for being cost-effective, saving time and energy, ease in doing the regular tasks, etc, such as the deepfake AI generated look-alike of Anjana Om Kashyap, a senior news anchor of Aaj Tak, used for tasks like reading headlines, multilingual bulletins, updating the public without requiring the physical presence[14]. Whereas on the other side, deepfakes are criticised for being a threat, spreading false information, misleading the public, etc., and it is here where the question is raised as to their significance and misuse.

On the one side, deepfakes are considered a significant advancement in several sectors such as the entertainment industry, education, and many other benefits, such as:

- **Caller Response Services-**

  In such services, the deepfakes are considered significant as they are used to provide personalised responses to the caller requests, which involve call forwarding and other receptionist services. For instance, telecom companies use deepfakes to offer recharge plans through recorded calls. Another instance is often seen when a huge lump sum amount is debited from the bank account, it is followed by a bank automated call to confirm a suspicious transaction, or to confirm whether such a transaction was valid.

- **Customer Phone Services-**

  This is the system wherein businesses use deepfakes in order to interact with customers by way of phone calls, providing support, information, or services such as complaint resolution, order-tracking, etc. Developers use fake voice, i.e., AI-generated fake

[14] Shweta Sharma, 'AI Anchors and the Future of News Broadcasting in India' *Indian Express* (2023)

voices, for simple tasks such as checking an account balance, or filing a complaint, etc. This ensures a quick response to the issues proving assistance, enhancing customer satisfaction, as many times the issue is already solved, and as a result, it assists the company as well as the customer, both saving from the hustle and trouble.

- **Entertainment-**

  In creating the enhanced visual effects, minimising production cost, and assisting in the script-writing, creative plots, and others, the deepfake secures significant role in the entertainment industry. It can be seen in Hollywood movies where the production houses use this for scenes that are hard to shoot, satire and parody content, in which the audience is aware that the videos are not real, and still enjoys the humorous situation created by the deepfake. This is further used for de-aging the actors and matching the lip movements to different languages to improve the dubbing, assisting the makers in creating seamless, immersive experiences while saving time and resources, making the production more efficient and visually compelling.

- **Education-**

  Deepfakes offer an immersive and engaging learning experience by creating AI-generated avatars that can pretend to be real historical figures, which allows interactive learning lessons for the students in a more dynamic manner. It helps in the better understanding of the concepts, as deepfakes can provide enhanced visuals of such concept, especially in the subject of science wherein the visuals give a better understanding of the medical, quantum of physics, etc. furthermore, it helps in retaining focus and facilitate leaning. It further transforms the traditional education into a more participative, engaging and inclusive experience.

However, as everything has another flip side, deepfakes are not an exception. Thus, the misuse of such deepfake technology has overshadowed its benefits, as misuse is raising significant questions as to the privacy and personal dignity of an individual. The viral deepfake video of Rashmika Mandanna[15], in which her face was morphed onto an Instagram video without her consent, illustrates the grave misuse. This is not the only way or incidents of deepfakes; there

[15] Soumya Gupta, 'Rashmika Mandanna Deepfake Sparks Debate' *Indian Express* (November 2023)

are several other ways in which it is misused to manipulate the public, such as-

- **Blackmail and reputation harm-**

  As deepfake videos and audios appear to be more authentic and realistic, victims fall in the trap and struggle to prove their innocence. The targeted picture is put in an unlawful, inappropriate situation, such as engaging in sexual acts or taking drugs. Such AI-generated images or videos are frequently misused to extortion, reputation harm, revenge-seeking, cyberbullying, etc. additionally, raising the concerns in regards to the consent and digital exploitation.

- **Fraud-**

  The violation of right to privacy and personal dignity is a gateway to commission of several other offences, as such personal information, including images and videos, which are ingested in the deepfake technology understands them and creates the familiar faces and voices by using such deepfakes wherein a replica of family members, company executives, bank officials, or public servants are used to exploit trust believing to be in their favour, to get personal and sensitive data such as bank account credentials, credit card numbers, or company credentials.

- **Political manipulation-**

  In the political domain, especially during the elections, such deepfakes are commonly misused to showcase political figures doing an act or saying something, which were actually never done or said, however, AI-generated images or videos, audios are fabricated and widely spread to mislead the public, influence their perception and opinion towards a particular political party. Furthermore, used to demonstrate the superiority of one political leader while making other inferior, for instance, during the Russia- Ukraine war, a deepfake video of Ukrainian President Volodymyr Zelenskyy announcing surrender to the Russian Army.

1.5 CONCLUSION

Deepfake technology characterized as one of the most complex and ironic advancements of the digital age. Even though it was an innovation with the main aim to enhance productivity, for

the ease of doing tasks, assisting in the management, and many other benefits, its capacity for manipulation and misuse cannot be overlooked, as they are not merely a technological advancement tool, but is an instrument that can manipulate and influence reality, violate privacy, and personal dignity, and harm reputation. Instances of deepfake videos of public figures like Rashmika Mandanna, Donald Trump, and many others, often circulated, shows its misuse and the impact it causes to a person's reputation and violating their right to privacy and personal dignity, and autonomy. Therefore, the innovations and technological advancements are not the issue or problem, but the lacunae that are left unnoticed or rather ignored to address attract the serious concerns, detailed examination of the existing laws, because society is dynamic in nature, it is bound to change, and at the same time, the legislature has the responsibility that laws are accommodating such changes and the society is well protected. The benefits of any innovation, including deepfakes, should be acknowledged, but they should also be accompanied by strong legal safeguards and regulatory mechanisms.