**Journal of Positive School Psychology**
2023, Vol. 7, Issue 1
Pp 225-230

# Deepfake: life & law in the age of illusions & manipulations

**Muhammad Sohail Asghar**
**Hamid Mukhtar**
**Kashif Mahmood Saqib**
*Assistant Professor of Law, University of Okara, Punjab, Pakistan*
**Hafsa Naz**
*(Corresponding Author) Visiting Lecturer, University of Okara, Punjab, Pakistan*

**Abstract**
Landscape of digital media and content creation is in continuous transaction on the advent of deepfake technology. Deepfakes, fueled by complex and sophisticated Artificial Intelligent (AI) algorithms, have the ability to manipulate audio-visual and photographic content in order to create hyper-realistic simulations of individuals and events. These technological advancements have opened up the new avenues of creativity on the one hand and ushered complex legal and ethical dilemmas on the other. Pakistan, a developing country with the rapid expansion of digital infrastructure and insufficient cyberlaw framework is under serious threat owing to proliferation of deepfake technology. Though, at the moment this technology has not gotten a foothold in Pakistan but by analyzing the heavy reliance on ICT nationwide, the same finally is unavoidable. This paper aims to explore the multidimensional legal and social implications posed by the deepfakes and concisely but lucidly examines the possible legal and regulatory strategies that may be opted in order to tackle the challenges arising from its use.

**Keywords:** Deepfake, Manipulated content, AI ethics, Cyberlaw, Pakistan

**Article History:**
Received: 24th Feb. 2023
Accepted: 28th Mar. 2023
Published: 19th Apr. 2023

## 1. Introduction:

A quick dive into the history reveals that our society is reshaped by many changes, some induced by the technology and others by politics or others elements. Deepfake, as a term, was coined in 2017 (Barber, 2019). It is used to refer a highly sophisticated artificial intelligent algorithm, capable to manipulate audios, videos and images to represent a scenario or incident. For example; to create a fake video where some specific individual is saying something or doing some certain act which he actually never said or did. Since its inception, this term has assumed a negative connotation mainly due to the risks that it represents to those who maybe its subjects and as well as those who are at the receiving end (Chesney & Citron, 2019). Deepfake could be disseminated rapidly and it can cause misinformation at a mass level. The deepfake video of ex-president US Barak Obama calling ex-president Trump a "complete dipshit" (Peele, 2018), or the claim of Mark Zuckerberg regarding having "total control of billions of people's stolen data" (Cole, 2019) are the evidence that how convenient it has become to spread a hyper realistically designed illusion. The term deepfake is amalgamation of two word i.e., "deep learning" and "fake". As the term itself suggests that deepfakes depend upon the neural networks to analyze very large sets of data samples in order to learn to copy someone's facial expression, accent, voice and gestures (Westerlund, 2019).

Generative Adversarial Networks or commonly known as "GANs" is the is a deep learning technique which is employed to create deepfake content. In this technique two deep neural networks known as "generator or actor" and "discriminator or critic" are trained in tandem (Mesky & Liaudanskas, 2020). Generator learns the statistical patterns of the sample data and generates a manipulated output. The other neural network i.e., discriminator attempts to distinguish between the real and the synthetic data. In this process, the generator attempts to trick the critic into believing that the manipulated data is the original one. The output is a reiterative progression whereby the feedback from the discriminator enables the generator to create hyper realistic deepfakes (Knight, 2018). In essence, the two neural networks act as picture forger and art detective (Giles, 2018), both in continuous struggle to outwit each other until it becomes impossible for the

art detective to identify the manipulation. Deepfake audio clips are also generated by using the same technology (Lyrebird, 2017). The deepfakes are becoming increasingly real and accessible with every passing day, Samsung has developed a highly sophisticated deepfake software which is commercially available (Solsman, 2019). Easy access of deepfake through multiple commercial services (Hill & White, 2020) pose serious threats at a societal and national level (Silbey, & Hartzog, 2019). While conspiracy theories are not new, misinformation alongwith weaponized deepfake represents a more complex situation to deal while combating challenges modeled by false information. A widely circulated deepfake has the capacity to deform democratic values, manipulate electoral process, jeopardize national security and can even provoke an armed conflict (Chesney & Citron, 2019).

Such conflict was occurred in 2019 when a supposedly deepfake of Gabonese president Ali Bongo created doubts regarding his health and resulted in an attempted coup (Breland, 2019).

## 2. How Deepfakes are used

It is a common saying that "effectiveness of the tool depends on the skillfulness of the hand that wields it." The same phenomenon could be applied on the deepfakes. Whilst deepfakes are unparallel when it comes to creativity and they carry unprecedented commercial value, there are numerous examples where they have been used in illegal ways to harm some specific individual or society at large. This paper categorizes the uses of deepfakes into two broad categories (i) Criminal & (ii) Commercial.

### i. Criminal Use

(a) The first ever use of GANs was to create deepfake adult content; particularly nonconsensual face-swapped still images of the celebrities where their faces were superimposed on the bodies of adult movie stars, later on deepfakes were used to create revenge porn where non-consensual adult videos were created and distributed by hackers or someone else seeking for the financial gain or to damage the reputation of someone after losing a romantic relationship (Mesky & Liaudanskas, 2020). Though, earliest deepfake adult content featured famous Hollywood actresses such as Scarlett Johansson and Gal Gadot but the harsh truth is that anyone could have superimposed his/her face on the body of some adult actor. Hence, deepfake adult content raises multiple legal concerns ranging from privacy breach to violation of intellectual property rights, consequently the victim suffers from humiliation, physical, psychological or financial abuse and irreparable loss of reputation etc.

(b) Deepfakes are also used as an effective instrument to manipulate public opinion or to shake trust in public institutions. The damaging effect of deepfake videos/photos/news reports could potentially affect the political career of a politician, for example deepfakes of a politician offering / receiving bribe, giving a racist speech or making confession of a crime could put an end to his political career (Siekierski, 2019). Deepfakes, weaponized by a hostile state pose even a greater threat to national security and can potentially impair the foreign relations of a state (Mesky & Liaudanskas, 2020).

The abovementioned examples of deepfake have the capacity to cause riots, domestic unrest and disturbance in elections. Reshaped policies of international community, regarding some particular state, based on the misinformation can ignite international conflicts destroying the lives of millions (Evans, 2018).

(c) Deepfakes pose a major challenge to cyberspace security. The recent time has witnessed multiple financial scams committed by the cybercriminals by using deepfakes. Deepfake technology can be deployed for the manipulation of stock or market; real time audio-visual representations are created of business executives making announcement for the launch of a fake product or concocted acquisition/merger, announcing bankruptcy or financial loses or making an order to an employee to make an urgent cash transfer; the list is infinite (Sullivan, 2019). With the rapid improvement of deepfake technology business entities need to adopt protective measures to spot them, if not they pose threat of brand sabotage, embarrass business management and finally result in forfeiture of business.

**ii. Commercial Use**

Like all other modern inventions, deepfake technology has also positive uses in several industries such as healthcare, education, entertainment, media, communication, fashion and e-commerce. Any business interested in the creation of distinct brand can utilize deepfake in order to substantially transform its products or services, to make distinguishable for its customer. For example, creation of an avatar for the customer of an e-fashion store to try on attires before making purchase (Baron, 2019). The benefits of deepfake technology are undeniable when it comes to the movie industry. Visual story telling has always remained a quite expensive experience for the directors and writers, now the deepfake technology has bestowed upon us the ability to merge images with the storyline at a fairly low cost in order to bring imaginative creativity to life. Examples of the potential use of deepfake in the movie industry include but are not limited to the creation of new movies starring dead actors, rendition of AI created voices for the actors who have lost their voices due to some medical reason, making use of special facial effects at post production phase, recreation of some scenes instead of reshooting them etc. (Dhillon, 2019). Overcoming the language barrier during a video conference and simultaneous alteration in the facial movements to match the language, in order to make a batter eye contact and make everybody appear to be speaking the exact same language is another remarkable achievement of the deepfake technology (Ballantine, 2019). Malaria awareness campaign of 2019 featured David Beckham; he appeared multilingual during this campaign due to the blessing of deepfake technology (Beckham, 2019). If used constructively, deepfake technology is a means to simplify creative interactions. Deepfake technology carries infinite possibilities; its true potential is yet to be exploited.

## 3. How to combat harmful deepfakes

Content fabrication has acted as catalyst to manipulate the opinion of public in recent times, the creation of echo chambers and filter bubbles have encouraged to adopt measures to tackle these effects. There are four basic strategies that could be adopted to combat deepfakes; (i) legislative and regulatory framework (ii) education and training (iii) retraction of legal immunity of social media firms (iv) adoption of anti-deepfake technologies.

**i. Legislative and regulatory framework**

Keeping in view the possible uses and abuses of the deepfake as discussed earlier in this article, the role of the legislative and regulatory framework can't be overemphasized. We could not find any reference specifically regarding deepfakes in the contemporary civil and criminal law of Pakistan. Nevertheless, a careful analysis of cyber laws of Pakistan disclosed the possibility for the adoption of these laws to redress the legal implications arising out of the use of deepfake technology. The given laws could be relied upon as basic principles while combating the negative impacts of the deepfake technology in Pakistan:

*(a) Personal Data Protection Bill, 2023*

Section 5: Grounds for processing personal data

Section 6: Consent for personal data processing

Section 15: Processing of sensitive and critical personal data

Section 17: Compliance with the data access request

Section 25: Right to prevent processing likely to cause damage or distress

Section 26: Right to erasure

Section 29: Right to data portability and automated processing

Section 31(2): Critical Personal Data shall only be processed in a server(s) or digital infrastructure located within the territorial jurisdiction of Pakistan

*(b) Digital Pakistan Policy, 2021*

Section 1.1.1: Online privacy and personal data protection through data protection law.

Section 1.1.2: Protection of cloud data privacy by devising legal framework for cloud-based services.

*(c) The Prevention of Electronic Crimes Act, 2016*

Section 3: Unauthorized access to information system or data

Section 4: Unauthorized copying or transmission of data

Section 14: Unauthorized use of identity information

Section 18: Offences against dignity of a natural person

Section 19: Offences against modesty of a natural person and minor

**ii. Education and Training**

Education and training are of prime importance while combating deepfake. People yet have failed to reckon the threats posed by the deepfake technology. Awareness campaigns at a mass level are the dire need of the hour to make people realize the true potential of AI if used negatively. The regulators need to understand that unlike reality, videos cannot be trusted to comprehend the actual accounts of a happening. Digital literacy alongwith the critical thinking must be taught at educational institutions; such training will equip the new generation with the tools to spot manipulated videos and interact with each other more respectfully in the cyberspace. These skills need to be promoted among the older as well in order to maximize tech-savvy population capable to critically determine the reliability, authenticity and social context of a video in order to access the real intent behind it. We need to acknowledge the fact that with the development of the technology, fewer photos of the faces will be needed to create a hyper realistic deepfake and that no one is immune. If we have ever posted a selfie or a video on a social site, we already have exposed ourselves to the risk of being deepfaked (Clark, 2019). Our best survival strategy against deepfake is to keep our photos and videos away from internet. Having some obstructions, for example waving hand, covering a portion of our face in a photo or video could also provide a little protection against deepfake (Solsman, cnet.com, 2019).

**iii. Retraction of legal immunity of social media firms**

Social media networks are the greatest means to spread any false or manipulated material instantaneously and they enjoy broad legal immunity with reference to the user shared content. Such legal immunity must be retracted to make social media platforms more vigilant and responsible regarding the material that their users post. Social media firms need to devise and implement more ethical policies to discourage deepfake content so that their network could not be weaponized for disinformation by the anti-state elements. Though, at the moment, most of the social media networks don't have any specific policies regarding deepfake but it is quite encouraging that these companies are gradually heading towards more sensible approach while processing user shared content. Social media networks such as Facebook and Instagram, and news agencies such as Wall Street Journal and Reuters cuts down any content which is identified as false, misleading or fake by third party fact checkers (O'Sullivan, 2019).

**iv. Adoption of anti-deepfake technologies**

Anti-deepfake technology provide us the most versatile tools to delete deepfake, check the authenticity of the content and prevent the existing content from being manipulated. Media forensics have described many indicators in order to detect deepfakes. Imperfections such as inconsistent gestures with the facial expressions, face wobble, inconsistency between the mouth movement and speech, irregular movement of the fixed objects such as the stand of the microphone, strange eyes movement and too smooth skin etc. are the obvious parameters to detect a manipulated video or image. Most likely targets of deepfake such as celebrities and

politicians can protect themselves from being deepfaked by wearing specially designed 3D printed glasses (Murphy, 2019). Advanced anti-deepfake technologies are based on the mammalian auditory systems, for example a mice can detect such inconsistencies in an audio that are either undetectable or ignored by humans (Mack, 2019). AI can verify the authenticity of a video in two ways, it can either check the full video frame by frame to detect any forgery or sign of manipulation or can examine it at once to access any inconsistencies as are mentioned above (Carbone, 2019). Blockchain technology is of utmost importance while deciding the authenticity and origin of a video as with every transfer it creates and stores digital signature in a ledger that is nearly impossible to manipulate. Social media platforms and news media companies can play vital role in the promotion of legitimate and authentic videos while discouraging the manipulated ones by implementing sophisticated AI filters to avoid any manipulated content. Another way to protect our data from being deepfaked is the insertion of noise into out videos and photos. Human eyes are imperceptible to this noise but it makes our visual material useless for the deepfake software (Walsh, 2019).

## 4. Conclusion

Positive uses of the deepfake technology can't be denied, so outright ban on this technology will be imprudent. However, threats that this technology poses are also serious as the deepfakes spread instantaneously reaching out to millions of people in dozens of jurisdictions within a few seconds. Legislation of a dedicated AI policy at the national level with a special focus on the deepfake technologies is the dire need of the hour. Such policy must not only define deepfake rather be capable to anticipate technological advancements while ensuring that human rights remain at the heart of such laws and policies. The most effective way to defeat the challenges introduced by the deepfakes involves an amalgamation of legal, educational and technical advances accompanied by an efficient enforcement mechanism.

## References

1. Barber, G. (2019) Deepfakes are getting better, but they're still easy to spot. https://www.wired.com/story/deepfakes-getting-better-theyre-easy-spot/?verso=true
2. Breland, A. (2019) The Bizarre and Terrifying Case of the "Deepfake" Video that Helped Bring an African Nation to the Brink. https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/
3. Baron, K. (2019) Digital Double: The Deepfake Tech Nourishing New Wave Retail https://www.forbes.com/sites/katiebaron/2019/07/29/digital-doubles-the-deepfake-tech-nourishing-new-wave-retail/?sh=6a8e46bb4cc7
4. Ballantine, M. (2019) Are deepfakes invading the office? https://www.forbes.com/sites/mattballantine/2019/07/03/are-deepfakes-invading-the-office/#19bf48923ea1
5. Beckham, D. (2019), David Beckham 'speaks' nine languages in call to end malaria, Global News, https://www.youtube.com/watch?v=U-mg7a1vwkw
6. Clark, C. (2019) Mona Lisa and Nancy Pelosi: The implications of deepfakes https://www.forbes.com/sites/charlestowersclark/2019/05/31/mona-lisa-and-nancy-pelosi-the-implications-of-deepfakes/?sh=4eca6a564357
7. Chesney, R. Citron D, (2019) Deep Fakes: A Looing Challenge of for Privacy, Democracy, and National Security, 107 California Law Review Pp. 1753-1819
8. Cole, S. (2019) a deepfake of Facebook founder Mark Zuckerberg surfaced online, wherein he appeared to seemingly give a sinister speech about Facebook's power to control billions of people's data. Available at https://www.vice.com/en/article/ywyxex/deepfake-of-mark-zuckerberg-facebook-fake-video-policy
9. Carbone, C. (2019) New tool detects deepfake with 96 percent accuracy, researchers say https://www.foxnews.com/tech/new-tool-detects-deepfakes-accuracy
10. Dhillon, S. (2019) The Optimistic view of deepfakes https://techcrunch.com/2019/07/04/an-optimistic-view-of-deepfakes/
11. Evans, C. (2018) 'Spotting Fakes News in a World with manipulated video https://www.cbsnews.com/news/spotting-fake-news-in-a-world-with-manipulated-video/
12. Giles, M. (2018) The GANfather: The Man Who's Given Machines the Gift of Imagination, MIT Tch. Rev. (Feb. 21, 2018)

13. Hill, K. & White, J. (2020) Designed to Deceive: Do These People Look Real to You?, N.Y. TIMES https://www.nytimes.com/interactive/2020/11/21/science/artificial-intelligence-fake-people-faces.html.
14. Knight, W. (2018) The US Military Is Funding an Effort to Catch Deepfakes and Other AI Trickery, MIT Tch. Rev. (May 23, 2018)
15. Lyrebird AI (2017), Lyrebird AI developed "voice-cloning" technology and released fake audio clips, including a clip of President Donald Trump discussing sanctions against North Korea. https://twitter.com/LyrebirdAi/status/904595326929174528
16. Westerlund, M. (2019) The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review. Vol 9, Issue 11
17. Mesky, E. Liaudanskas, A. et al. (2020) "Regulating Deepfakes: Legal & Ethical Considerations", Journal of Intellectual Property Law & Practice, Vol 15, No. 1 p. 26
18. Mack, E. (2019) Researchers propose detecting deepfakes with surprising new tool: Mice https://www.cnet.com/science/researchers-propose-detecting-deepfakes-with-surprising-new-tool-mice/
19. Murphy, H. (2019) Cyber security companies race to combat deepfake technology https://www.ft.com/content/63cd4010-bfce-11e9-b350-db00d509634e
20. O'Sullivan, D. (2019) The Democratic Party deepfaked its own chairman to highlight 2020 concerns https://edition.cnn.com/2019/08/09/tech/deepfake-tom-perez-dnc defcon/index.html
21. Peele J. (2018) Video of Ex-President Barack Obama giving a public-service announcement about the dangers of deepfakes, an announcement that Obama actually never made. Available at https://www.youtube.com/watch?v=cQ54GDm1eL0
22. Solsman, J. E. Samsung deepfake AI could fabricate a video of you from a single profile pic, CNET https://www.cnet.com/news/samsung-ai-deepfake-can-fabricate-a-video-of-you-from-a-single-photo-mona-lisa-cheapfake-dumbfake/
23. Silbey, J. & Hartzog, W. (2019) The Upside of Deep Fakes, 78 Maryland Law Review 960-66
24. Siekierski, B.J. (2019) Deep Fakes: What can be done about synthetic audio and video? https://lop.parl.ca/sites/PublicWebsite/default/en_CA/ResearchPublications/201911E
25. Sullivan, D. O. (2019) House Intel chair sounds alarm in Congress' first hearing on deepfake videos https://edition.cnn.com/2019/06/13/tech/deepfake-congress-hearing/index.html
26. Solsman, J. E. (2019) Deepfakes may ruin the world. And they can come for you, too https://www.cnet.com/tech/computing/deepfakes-may-try-to-ruin-the-world-but-they-can-come-for-you-too/
27. Walsh, C. (2019) What is a deepfake? This video technology is spooking some politicians https://www.usatoday.com/story/news/politics/2019/03/15/what-deepfake-video-technology-spooking-some-politicians/3109263002/