

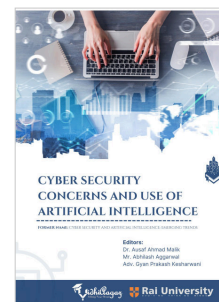
## CHAPTER 8

# The Dark Mirror of AI: Combatting the Rise of Deepfake Technologies

Vijpreet Pal

in *Cyber Security Concerns and Use of Artificial Intelligence*, edited by Ausaf Ahmad Malik, Abhilash Aggarwal and Gyan Prakash Kesharwani (VidhiAagaz 2024), pp. 74–79

ISBN 978-81-947620-5-8



## CITATIONS

BLUEBOOK (21ST ED.)

Vijpreet Pal, *The Dark Mirror of AI: Combatting the Rise of Deepfake Technologies*, in *CYBER SECURITY CONCERNS AND USE OF ARTIFICIAL INTELLIGENCE* 74 (Ausaf Ahmad Malik, Abhilash Aggarwal & Gyan Prakash Kesharwani eds., VidhiAagaz 2024).

OSCOLA (4TH ED.)

Vijpreet Pal, 'The Dark Mirror of AI: Combatting the Rise of Deepfake Technologies' in Ausaf Ahmad Malik, Abhilash Aggarwal and Gyan Prakash Kesharwani (eds), *Cyber Security Concerns and Use of Artificial Intelligence* (VidhiAagaz 2024) 74.

CHICAGO (17TH ED.)

Pal, Vijpreet. "The Dark Mirror of AI: Combatting the Rise of Deepfake Technologies." In *Cyber Security Concerns and Use of Artificial Intelligence*, edited by Ausaf Ahmad Malik, Abhilash Aggarwal and Gyan Prakash Kesharwani, 74–79. Maihar: VidhiAagaz, 2024. <https://vidhiaagaz.com/books/cyber-security-concerns-and-use-of-artificial-intelligence/chapter-8/>.

APA (7TH ED.)

Pal, V. (2024). The dark mirror of AI: Combatting the rise of deepfake technologies. In A. A. Malik & A. Aggarwal & G. P. Kesharwani (Eds.), *Cyber security concerns and use of artificial intelligence* (pp. 74–79). VidhiAagaz. <https://vidhiaagaz.com/books/cyber-security-concerns-and-use-of-artificial-intelligence/chapter-8/>

## SOURCE

<https://vidhiaagaz.com/books/cyber-security-concerns-and-use-of-artificial-intelligence/chapter-8/>

## RIGHTS

Free to read online. © VidhiAagaz 2024. All rights reserved: no part of this publication may be copied, reproduced, adapted or transmitted in any form or by any means without the prior written permission of the publisher.

Citations are given as a guide; check them against the style manual you follow. This page was added by VidhiAagaz to the downloaded file. The chapter follows, with the page numbers of the printed book.

# THE DARK MIRROR OF AI: COMBATTING THE RISE OF DEEPPFAKE TECHNOLOGIES

Vijpreet Pal<sup>9</sup>

---

## **ABSTRACT**

*Deepfakes are fabricated, manipulated content created by merging, altering, replacing, and superimposing images, videos, or audio using artificial intelligence and machine learning. This advanced technology produces remarkably realistic results, portraying individuals saying or doing things they never actually did. While deepfakes open avenues for creativity in fields like education and art, their misuse has surged, particularly in illicit and harmful ways. A large proportion of deepfake videos involve non-consensual pornography and revenge porn, posing severe threats to personal dignity and privacy. In the political arena, deepfakes hold the potential for weaponization, where fabricated content could incite military coups or justify conflict between nations. The increasing sophistication of synthetic media also raises concerns about the diminishing credibility of audio and video evidence, as these forms of media are increasingly used to facilitate disinformation, damage reputations, and infringe on privacy. This paper seeks to highlight the growing dangers posed by deepfakes as a result of advancements in artificial intelligence. It examines the technology behind deepfakes and discusses how these synthetic media can be used maliciously to harm individuals and society at large. The discussion covers various real-world instances of deepfakes being used to commit misdeeds, along with hypothetical examples illustrating their potential for further misuse. As philosopher Jean Baudrillard said, "The simulacrum is never that which conceals the truth—it is the truth which conceals that there is none." In an era where synthetic media blurs the line between reality and fabrication, the consequences are profound. Additionally, the paper explores current regulatory measures and proposes legal frameworks to mitigate the risks posed by deepfakes, offering potential solutions to address privacy breaches, reputational harm, and the broader societal impacts of this emerging technology.*

**Keywords:** deepfakes, artificial intelligence, data privacy, propaganda, defamation, copyrights.

---

With the use of deepfake technology, voices and faces may be added to audio and video recordings. The results are pretty accurate. The term "deepfake" was coined by a Reddit user in the year 2017, who used the technology to alter the faces in pornographic videos. Primarily deepfakes are employed in four cases and they are - pornography, political campaigns, commercial uses to reduce transaction cost and creative work like movies, memes, etc. Two videos featuring Bharatiya Janata Party candidate Manoj Tiwari condemning Delhi's Aam Aadmi Party were manipulated using "deepfake technology" to appear as he was speaking in English and Haryanvi, aiming to reach more voters during the 2020 legislative assembly election in Delhi. In 2019, Silversparro Technologies, a tech start-up engaged in machine learning and deep learning generated a deepfake GIF of Narendra Modi, prime minister of India on experimental basis. Abhinav Kumar Gupta, founder and CEO of Silversparro Technologies said that "[T]he implications of deepfakes are worrying" and claimed that a dystopian future awaited the Indian politics.

Shortly after the official launch of the technology in 2018, even websites like Twitter and Facebook have banned it on their platforms. An unprecedented rise in the generation of these deepfake videos has

---

<sup>9</sup> Student at United University, Prayagraj, India.

been observed, leading to probable legal consequence in the areas of defamation, free expression and privacy. Therefore, it is important to acknowledge that the existing legal provisions are neither sufficient to curtail the application of deepfake nor guarantee justice to the aggrieved parties. *Part II* of this article explores deepfake technology, while *Part III* investigates potential misuses. *Part IV* discusses Indian regulatory and legal remedies to combat deepfakes. *Part V* details legislative and judicial actions in the UK against deepfakes. *Part VI* concludes with conclusive remarks.

### **TECHNOLOGY BEHIND DEEPPFAKES**

Franklin Foer of the Atlantic described deepfakes as, “[O]ne of the cruelest, most invasive forms of identity theft invented in the internet era.” Deepfakes are primarily based on machine-learning algorithms which harness facial mapping software to produce a fabricated content of an individual's identity without their permission. The technology uses “neural networks” which, begins with a tabula rasa kind of a layout with a nodal network managed by a set of numerical standard controlled randomly. The neural networks are trained by examples, quite similar to how a human brain is trained by experiences. Likewise, the network is able to produce increasingly accurate impersonations or alterations by processing a broad set of training examples. Further, the generation of increasingly convincing deepfakes has shot up post the invention of “Generative Adversarial Networks” by Ian Goodfellow. The approach involves the co-existence of two neural networks concurrently, one known as the “generator” and the other known as the “discriminator”. The function of the generator is to draw on a dataset to generate a sample that would further mimic the dataset. The function of the discriminator is to assess the extent to which the generator has succeeded with the mimicking process. The assessments are informed by the discriminator to the generator in an iterative manner. After the completion of the preliminary machine learning process, the program can run without any human supervision.

Essentially, there exist three kinds of deepfake videos i.e. Face swapping which involves superpositioning synthesised faces on target faces while maintaining facial expression. Head puppetry synthesises target head and shoulder using source head, while lip syncing alters lip region to create fabricated video, resulting in unspoken speech.

### **POTENTIAL MISUSES OF DEEPPFAKES**

*“Oh, what a tangled web we weave...when first we practice to deceive.”* Technology can bring benefits to individuals and society, but it can also be misused. Deepfakes, which exploit advanced information systems, can cause serious harm. They can be used by individuals, corporations, and enemy states to create falsehoods and harass individuals. The Indian Constitution allows for reasonable restrictions on freedom of speech and expression.

#### ***(A) Threat to Security, Sovereignty of the State and Friendly Relations with Foreign State***

Deepfakes present a glaring threat to national security. A Federal Bureau of Investigation (United States security agency) executive in a summit recently held stated that: *“What we're concerned with is that, in the digital world we live in now, people will find ways to weaponise [sic] deep-learning systems”*.

India and Pakistan have been involved in three bloody wars due to false flags. Recently, Pakistan army created fake Twitter profiles of Indian generals to spread fake stories, justifying incursions. Deepfake technology presents a significant threat to the Indian government and army if enemy states use it to defame India or take aggressive actions.

### ***(B) Violation of Public Order and Incitement of Violence***

Public order which is derived from the French word *ordre publique* in normal parlance means the maintenance of not only the individuals' duties but also the peace and harmony of the society at large. Public order in India is being violated by mob lynchings and violent protests, largely due to viral WhatsApp messages spreading fake warnings of child-trafficking, organ harvesting, or witchcraft. The use of deepfakes technology is exacerbated, leading to a growing number of malicious attempts to create panic and disrupt society's peace.

### ***(C) Threat to Decency, Morality and Defamation***

The Supreme Court of India in the judgment of *Ranjit D. Udeshi v. State of Maharashtra* had stated pornography and libidinous characters in popular culture both qualify as obscenity. Deepfakes exploiting Indian movie stars are already out on the web. Noted actresses like Shraddha Kapoor and Deepika Padukone have been victimised by deepfake porn. The technology is being improved with an Android app, allowing users to create nude images of women in minutes.

India has been a witness to numerous cases of revenge porn. In 2014, a college girl in Kerala, India, was shocked to discover her private pictures were shared online by her former boyfriend. In another gruesome breach of privacy, a separated husband from Ahmadabad allegedly used social media to post private chats and naked pictures of his wife and sister-in-law, claiming them as prostitutes. Thus deepfakes present a massive threat to the decency and morality of the society.

### ***(D) Contempt of Court***

The criminal contempt often comes at loggerheads with the rights accorded to press and the freedom of expression. Supreme Court in *Harijai Singh, In re*, clarified by saying that a free and outspoken press is the driver of a healthy democracy but the said right is not absolute and unfettered at all times. In December 2019, the Karnataka High Court put out a notice of contempt of court against the daily newspaper Deccan Herald on account of a fake news publication. The report had vehemently stated that the High Court had issued notices to 48 Judges and that more than 20 Judges had taken voluntary retirement. Deepfakes could undermine the judiciary system by promoting fake reporting, doctored videos of bribes, and criminal contempt of courts, potentially gaining credibility.

## **TACKLING DEEPPAKES IN INDIA: ISSUES, SOLUTIONS AND REGULATORY STEPS**

### ***(A) Hindrances to Legal Redressal***

Deepfakes are often created on conditions of anonymity. Civil and criminal charges cannot be levied upon the creators of deepfakes if they are not identified. Law enforcers find it difficult to track Internet Protocol (IP) addresses of the pirates due to anonymous and secretive nature of their operations. Similarly, if a person uploads deepfake porn as an act to defame a person, he would do so after taking

all the necessary precautions to hide his IP address. The rise of Virtual Private Networks (VPN) has added to the problem. VPNs allow individuals to shift their IP address to remote servers through a tunnel in order to stay invisible.

Second legal issue lies in the territoriality. With the access to Internet being global, perpetrators from other jurisdictions could try to defame Indian citizens or try to disrupt the peace in the nation by using deepfakes. In such scenarios taking action against it would be beyond the jurisdiction of the Indian civil or criminal courts.

### ***(B) Potential Legal Recourses***

#### **i. Suits against Copyright and Celebrity Rights Infringement**

Deepfake videos often use copyrighted content without valid licenses, infringing on the copyrights of the rightful owner. If a deepfake uses protected photographs or videos, the owner can file a civil suit for injunctions and damages. Celebrities can also sue for infringement of their “celebrity rights” as recognized by the Bombay High Court in *Sonu Nigam v. Amrik Singh*. Similarly, to combat deepfake pornography and deepfake parody videos created to disrupt the image rights of a celebrity, relief can be sought from the courts.

#### **ii. Defamation suits**

Deepfake revenge porn created by a person with a sole intention to devalue and harm the reputation of the victim would qualify as defamation. This would allow the aggrieved parties to file a criminal complaint against the perpetrator. If the person who made the deepfake porn is found guilty, he could face a prison term of up to two years and a fine.

#### **iii. Filing for criminal remedies against pornography**

In India, several laws have levied a strict ban on publication of obscene material or any content containing indecent representation of women. Section 67 of the IT Act imposes criminal liability for publication or transmission of obscene materials in electronic form. A very relevant judgment was rendered by the Supreme Court in *Sharat Babu Digumarti v. Govt. (NCT of Delhi)* which ruled that “*an activity emanating from electronic form which may be obscene is exclusively punishable under Section 67 of IT Act and not under Section 292 of IPC.*” Furthermore, with the dawn of internet age the courts in India followed the observation made in *Ajay Goswami v. Union of India* according to which in case of a combination of art and obscenity it is imperative to determine if the “*artistic, literary or social merit of the work in question outweighs it obscene content.*”

Further, strict legal actions can be taken against *transmitters* of such videos especially if its child porn. The Protection of Children from Sexual Offences Act, 2012 is the primary legislation against child pornography, committing individuals involved in its preparation, production, offering, transmitting, publishing, facilitating, and distribution for sexual pleasure. Child pornography has also been banned explicitly under the IT Act. Thus a person using deepfakes to make child pornography can be punished as per these provisions.

#### **iv. Remedies for breach of data privacy**

In the case of *K.S. Puttaswamy v. Union of India*, the Supreme Court recognised the right to privacy as a fundamental right as per Article 21 and stated it to be the “constitutional core of human dignity.” Subsequent to the *Puttaswamy* decision, a chair headed by Justice Srikrishna had proposed the Personal Data Protection Bill, 2018 which was then tabled in the Lok Sabha as Personal Data Protection Bill, 2019. The Bill defines personal data as any data related to a natural person, including their characteristics or attributes. It defines a personal data breach as accidental disclosure or sharing of personal data that compromises confidentiality, integrity, or availability. Deepfake makers, who share personal data like videos or photos, are guilty of such breaches. Individuals exposed by these breaches can seek remedies under Section 25.

#### **DEEPPAKE REGULATION: UK PERSPECTIVE**

Currently, no European or national legislation exists in France, Germany, or the UK to address deepfakes. The European Commission has created a “Code of Practice on Disinformation” to address online disinformation. The UK Government’s “Online Harms White Paper” discusses deepfakes using artificial intelligence but does not explicitly address them. A recommendation for companies to improve user security and address online service issues remains a proposal.

The aggrieved party can use Section 1 of the Defamation Act, 2013 to prove that a post online caused or may cause grave reputational harm. They can also prosecute an Internet Service Provider (ISP) as they regulate hosts' actions on their platforms. The case of *Payam Tamiz v. Google Inc.*, was the first in UK wherein the liability of ISP was considered for libel. In the landmark decision rendered by the European Court of Justice in *Google Spain SL v. Agencia Española de Protección de Datos*, it was ruled that those individuals bearing the brunt of prejudicial search engine results have the “right to be forgotten” and pronounced that all the search engine operators are “Data Controllers” under the European Data Protection Law. The decision ruled that the processing of personal data of an individual may be disabled when the search is “[I]nadequate, irrelevant, no longer relevant or excessive”. However, “right to be forgotten” cannot be exercised by non-living entities such as a company. Additionally, if the “right to be forgotten” is not extended universally then the adverse Google search results will show up in other countries, thereby defeating the objective behind the ruling.

The Data Protection, Privacy and Electronic Communications (Amendment etc) (EU Exit) Regulations 2019 (DPPEC) define personal data as any information related to an identifiable natural person. The UK’s DPPEC is a revised version of the EU’s GDPR. Deepfakes technology, such as face swapping and lip-syncing, can incorporate personal data from the target, but head puppetry, which combines face, voice, body, or videos, cannot be considered personal data. However, constructive use of a person’s video or audio requires consent, and the relevant person has the “right to erasure” or “rectification”. The concept of “lawful processing of personal data” under DPPEC or EU GDPR is relevant to assess the legality of creating fake content. A data subject is vested with the “right to object” if any of their personal data is not processed in compliance to Article 6 of DPPEC or EU GDPR which provides for “lawful processing”.

Additionally, other potential measures could be fixation of automated fake content detection software in search engines and social media platforms and embedding inaudible “noise” to digital images which would make it difficult for the deepfake technology to process thereby affecting the quality of the deepfake produced. The UK’s existing remedies for addressing deepfake content are ineffective due to challenges in establishing claims due to identity issues and jurisdictional barriers, highlighting the need for improved legislative frameworks to include image rights, online content registration, and penalties.

## **CONCLUSION**

Lies in varied forms have been a part of the society since the beginning and to be realistic it is here to stay. Lies are often used to make jokes, as a method of parenting or as means to boast. But they have a capability to cause significant harm to an individual or the society at large. On that note, deepfakes will merely give a modern technical outlook on the art of lying. The main problem with deepfakes is with its potentiality of misuse by miscreants. Its cutting edge AI mechanisms can be taken use of by criminals to commit atrocities like defamation, national security breach, fake revenge porn, public disorder and contempt of court.

While the field of law around deepfakes in India is evolving, there is no specific statute or legislation for its regulation as of now. So, the task is up to the Indian Parliament to formulate a comprehensive body of law in case the threat of deepfakes rises exponentially. But the threat though mild exists currently. For the time being, the individuals that are victimised by it can use multitudes of civil and criminal remedies to get necessary justice.