DEEP FAKES AND PRIVACY: LEGAL CHALLENGES AND EMERGING SAFEGUARDS

Yashwant Shailendra Soni, Mahatma Gandhi Kashi Vidyapeeth

ABSTRACT

An artificial intelligence-based technique known as "deep-fake technology" creates or modifies audio-visual information while raising serious privacy issues. Deep-fake technology allows very lifelike representations of people in inappropriate circumstances while causing negative effects through election interference and disinformation efforts that impact both society and individual reputations. The General Data Protection Regulation (GDPR) of the European Union, certain state privacy laws in the United States, and Indian data privacy laws were all constructed without taking into account the unique risks associated with material produced by artificial intelligence. Deep-fakes can now take advantage of substantial legal issues pertaining to consent and data protection rights over personal photographs that have been created by legislative deficiencies. This study examines deep-fake technology and the serious privacy risks it poses, assessing technological and legal solutions. From a privacy perspective, this study looks into how deepf-fakes challenge conventional wisdom regarding consent, authenticity, and individual identity. From a technology perspective, the study looks at current deep-fake detection and prevention systems. It takes intricate procedures that call for both technological advancements and supportive legislative frameworks to successfully integrate these technologies into all-encompassing privacy protection frameworks.

Keywords: Keywords: Deep-fakes, Privacy, AI Regulation, Data Protection, Consent Etc.

Introduction:

A number of disruptive technical phenomena have emerged in the first quarter of the twenty-first century, such as the "millennium bug," the widespread use of ICTs, and the explosive growth of social media. Additional advancements include the weaponization of the internet and the revolutions in artificial intelligence (AI) technologies. The latter phenomena typically result in chaos with dire repercussions in the real world and poses a serious threat to the cyber resources and assets of businesses and countries as well as the security and welfare of internet users. Deep-fake software tools and mobile applications that are openly accessible have made it easier for anybody to create and use deep-fakes for both harmful and creative purposes. Deep-fakes are a subset of synthetic media that employ machine learning and artificial intelligence to produce realistic, convincing movies, photos, audio, and text of unreal events, igniting discussions about ethics, society, and the law. [1]

A "deep-fake" is defined as "AI generated or manipulated image, audio, or video content that looks like existing persons, objects, places, or other entities or events and is intended to be believed by a person as real or true" by the European Union's recently passed Artificial Intelligence Act (EU AI Act). In essence, deep-fakes are a type of deep digital manipulation—yes, as complex as AI-powered digital photo shopping.

Videos, pictures, and audio snippets that show individuals saying things or acting in ways that didn't happen in real life are known as synthetic media content, and its veracity depends on how the makers portray themselves. While some uses of synthetic media, such as movies, video games, and instructional materials, are safe and meant to be entertaining, other uses are extremely dangerous. Deep-fakes have become more common as artificial intelligence (AI), more especially machine learning (ML) and deep neural networks (DNNs), has developed. Even basic deep-fakes have the ability to disseminate misinformation quite well because of the human propensity to always believe visual evidence, which makes the technology itself seem like a tiny concern.

Furthermore, in order to deceive the public in political situations, politicians' faces are being altered to appear on other bodies in ways that, for example, inaccurately portray them making statements. Additionally, hackers utilise deep-fakes to pose as CEOs and other business

[1] **Sara Brown,** *Deepfakes Explained: What They Are and How to Spot Them*, MIT SLOAN SCH. MGMT. (Feb. 14, 2023), https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained.

officials in an effort to fool staff members particularly those in the finance department—into sending money to accounts controlled by scammers. [2]Although deep-fakes are frequently used for amusement, it has been demonstrated that cybercriminals also utilise them for fraud and disinformation. This study looks at how deep-fakes endanger people's right to privacy globally. It discusses the methods used to produce deep-fakes as well as their benefits and drawbacks. It also makes recommendations for mitigating the effects of potential hazards and preventing deep-fakes.

The Concept of Deep-Fakes:

Deep-fake technologies (DT) are now a serious threat to international organisations due to the development of artificial intelligence (AI). Deepfake is a kind of artificial intelligence-generated content that manipulates multimedia to mimic real-world events using deep learning techniques[3]. A growing number of altered films featuring people saying or doing things they never did and rejoicing in locations they have never been to are making the rounds on the Internet. The phrase "deep-fake" refers to the idea that deep learning techniques, which are a particular subset of machine learning, can be used to automatically identify and categorise intricate data patterns. The phenomenon of deep-fakes, which involves modifying films to attract a lot of attention, is widespread.

A phone clip of Barack Obama criticising President Donald Trump and another of Mark Zuckerberg making derogatory remarks regarding user data are examples of these kinds of videos. majority of Deep-fakes demonstrations have highlighted the possibility of fraudulent behaviours, incorrect information, and privacy problems. [4]

Deep-fakes and other emerging technologies raise significant concerns about how their growth may affect our social and cultural standards. Naturally, the growth of large picture databases, GPU and neural network technology, and other related advancements may be responsible for the surge of deep-fakes, but it is also clear that these technologies pose a risk. As this technology develops, it's crucial to consider not just how it will impact privacy issues

[2] **Oxford Internet Institute,** *Social Media Manipulation by Political Actors an Industrial-Scale Problem – Oxford Report*, University of Oxford (Jan. 13, 2021), https://www.ox.ac.uk/news/2021-01-13-social-media-manipulation-political-actors-industrial-scale-problem-oxford-report.
[3] **Bart van der Sloot & Yvette Wagensveld,** *Deepfakes: Regulatory Challenges for the Synthetic Society*, 46 **Comput. L. & Sec. Rev.** 105716 (2022), https://doi.org/10.1016/j.clsr.2022.105716 (last visited June. 20, 2025).
[4] **Douglas Harris,** *Deepfakes: False Pornography Is Here and the Law Cannot Protect You*, 17 **Duke L. & Tech. Rev.** 99 (2019).

in the modern era, but also how it will end and what impact it will have on the status of authenticity, truth, and trust in digital media.

Risks with Deep-fakes:

A helpful tool is only as good as the people using it, and deep-fake technology is no exception. Even though deep-fakes have the potential to be very innovative in both editorial and commercial contexts, there are numerous examples of deep-fakes that are purposefully produced to negatively impact society.

Cybersecurity and Deep-fakes: Deep-fakes pose a serious threat to financial fraud. Criminals are willing to fabricate recordings of business executives in order to influence markets, spread misleading information, or even issue orders for unlawful activity. This may result in legal problems, blackmail, harm to the company's reputation, and financial loss.[5]

Deep-fakes and Politics: Deep-fakes are a tool that political parties and actors can use to disseminate misinformation. They can worsen societal divide, damage reputations, fabricate tales, and influence elections by showcasing candidates' phoney claims. They can be used to undermine the country's international stability and national security, which can lead to mistrust of all information.

Deep-fakes and Sex-Related Content: Deep-fakes are first utilised to produce phoney sex videos. It involves making deep-fake revenge porn for financial gain and plastering the faces of celebrities on adult actors. All were carried out with the intent to hurt the victims and violate their right to privacy, image rights, and consent.[6]

Deep-fakes are therefore harmful since they pose a serious risk to individuals, organisations, and governments due to their ability to tamper with widespread disinformation. Deep-fake production is also getting better and more affordable. And society, business, and governments must fight this. The process includes developing new legislation and technological tools to control the production, ownership, and distribution of deep-fakes as well as enforcing penalties for their usage in any way.

---

[5] **Britt Paris & Joan Donovan,** *Deepfakes and Cheap Fakes*, Data & Society (Sept. 18, 2019), https://datasociety.net/library/deepfakes-and-cheap-fakes/.

[6] **Chidera Okolie,** *Artificial Intelligence-Altered Videos (Deep-fakes) and Data Privacy Concerns*, 25 **J. Int'l Women's Stud.** 13 (2023).

Legislative Frameworks:

A comprehensive legal framework that would assist in addressing those issues is becoming more and more necessary as deep-fake technology continues to advance. The current legal frameworks are unable to keep up with the exponential growth of generative AI, which leads to the spread of false information, invasions of privacy, and other negative effects.

The European Union's approach to AI Regulation:

The Artificial Intelligence Act of the EU:

The EU's more comprehensive regulatory approach to artificial intelligence includes deep-fakes. In order to support reliable and safe AI applications that are in line with the basic rights and values of EU citizens, the European Commission proposed a single AI law in April 2021. It lays out uniform guidelines for the creation, promotion, and application of AI systems.

AI systems are categorised in a risk-based scheme as posing unacceptable or high, limited, or minimal risk, in accordance with this suggested AI regulatory framework. In this sense, the Commission wants to outlaw AI systems since they pose an intolerable risk to the security and fundamental rights of EU residents. AI systems that are deemed "high risk" would be subject to requirements including risk assessments, documentation, human supervision, and the usage of high-quality datasets. Certain AI systems would be exempt from regulation, while others that present "minimal risk" would be subject to less strict rules.

The proposed rule requires modified content to be labelled, specifically targeting deep-fakes. Users who create or alter image, audio, or video content that effectively mimics real-world objects and could be confused for authentic (also known as "deep-fakes") are required under Article 52(3) to declare that the content is artificial. However, there are exceptions to this labelling need for the arts, sciences, and the practice of free speech, as well as for law enforcement activities (detection, prevention, investigation, and prosecution of criminal offences). Although the labelling requirement is a first step in reducing the possible risks of deep-fakes, it is unclear how it will be implemented in practice. [7]

First of all, there are no clear standards in the plan about the structure and content of these

[7] **Regulation (EU) 2024/1689, art. 52(3)** (European Parliament & Council June 13, 2024).

disclosures, which raises concerns about how they should look and how deep-fake technology vendors might help with labelling. Second, there are no clear sanctions in place to punish users who disregard Article 52(3)'s responsibilities for transparency. Non-compliance with the deep-fake labelling requirement is not specifically addressed in Article 71, which outlines sanctions. Lastly, it is doubtful if this strategy will be effective against malevolent actors that commonly distribute deep-fakes anonymously because they can easily avoid detection and, consequently, the labelling requirement. [8]

The GDPR:

The data needed to create deep-fakes as well as, often, the deep-fake itself are both covered by the General Data Protection Regulation (GDPR). As it relates to that particular person, the portrayal of someone carrying out actions they did not take qualifies as personal data.xii Regardless of the accuracy of the data, the GDPR's applicability depends on an individual's identifiability. Therefore, the main factor is the reasonable identifiability of the person portrayed; the altered nature of video or audio output produced with GANs or other technologies is irrelevant.[9]

The GDPR requires that all processing of personal data have a legitimate basis. "Informed consent" and "legitimate interests" are the most likely qualifying grounds in the context of deep-fakes. The rights and liberties of the person portrayed may take precedence over the legitimate interest of the deep-fake author, which could be the case for satirical or parodic deep-fakes that invoke the right to free expression. Deep-fake production and distribution necessitate the informed agreement of all identifiable individuals in both the original and manufactured content, absent a valid interest. A GDPR breach occurs when consent is not obtained. Although Member State legislation may demand the heir's consent, these consent requirements do not apply to deceased people.

The GDPR gives victims the right to have erroneous data corrected or erased, providing channels for combating illegal deep-fakes.

All Member States are subject to independent supervisory authorities that ensure GDPR

[8] **Regulation (EU) 2024/1689, art. 71** (European Parliament & Council June 13, 2024).
[9] **Bart van der Sloot & Yvette Wagensveld,** *Deepfakes: Regulatory Challenges for the Synthetic Society*, 46 **Computer L. & Sec. Rev.** 105716 (2022), https://doi.org/10.1016/j.clsr.2022.105716.

compliance. However, because of the anonymity of the offenders and the possible lack of finances for the victims, seeking legal action can be difficult. The right to manage one's image is acknowledged by the European Court of Human Rights as essential to one's personal growth. The "household exemption," which prohibits processing personal data for purposes that are solely domestic or personal and have no business or professional affiliation, is a major restriction of the GDPR. Although this exemption was created for private correspondence and networking, it is debatable whether it still applies in the modern data-driven world where the lines between the public and private domains are blurred.[10]

India's Legal Framework and Deep-fake Threats:

Although there is presently no specific law in India that addresses deep-fake threats, there are possible channels for recourse through existing statutes including the Information Technology Act, 2000 (IT Act), the BNS (2023) and the IT Rules, 2021.A directive requiring adherence to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (IT Rules) was released by the Ministry of Electronics and Information Technology (MeitY) on December 23, 2023.[11] The growing frequency of deep-fakes and AI-driven disinformation triggered this directive, which mandated that intermediaries notify users explicitly about content that is forbidden by Rule 3(1)(b) of the IT Rules, including in terms of service, privacy policies, and user agreements.

The IT Act of 2000 and the IT Rules of 2021 both clearly require social media companies to take down deep-fake images or videos as soon as possible. A fine of Rs 1 lakh or up to three years in prison are possible punishments for noncompliance. Anyone found guilty of false impersonation on a communication device or computer resources faces up to three years in prison and a fine of up to Rs 1 lakh under section 66D of the IT Act, 2000. [12]

In order to stop users from uploading content that impersonates someone else, they must comply with Rule 3(1)(b)(vii); additionally, Rule 3(2)(b) requires them to remove the content

[10] **ManageEngine,** *The GDPR Articles Explained: Chapter 6 – Independent Supervisory Authorities*, ManageEngine (last updated Apr. 1, 2024),

[11] **Ministry of Electronics and Information Technology (MeitY),** *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021*, § [as updated regional reference] (last updated Apr. 6, 2023),https://www.meity.gov.in/static/uploads/2024/02/Information-Technology-Intermediary-Guidelines-and-Digital-Media-Ethics-Code-Rules-2021-updated-06.04.2023-1-2.pdf (last visited June 14, 2025).

[12] **Information Technology Act, 2000**, § 66D, No. 21, Acts of Parliament, 2000 (India).

within 24 hours of receiving a complaint[13].

Controlling Deep-fake Pornography:

The exponential rise in non-consensual pornography production, which disproportionately affects women, is one of the most heinous effects of deep-fake technology. Section 67A of the Information Technology Act (IT Act), which makes it illegal to publish or transmit overtly sexual content, is an attempt to address this issue in India's legal system. This offence carries serious consequences; a first conviction carries a maximum term of five years in prison and a fine of ten lakh rupees, while successive convictions carry even more severe penalties. Furthermore, Section 67 of the IT Act covers content deemed obscene but not specifically pornographic, with harsher penalties for first-time offenders but nonetheless significant consequences, particularly for repeat offenders. [14] This demonstrates how deep-fake technology has enabled the legal assessment of the harm caused by both suggestive and explicit kinds of internet sexual content.

Technology-Based safeguards:

In this instance, deep-fake technology is developing, and strong tools and resources are becoming more and more necessary to counteract misleading information and raise awareness of it. People and organisations need to arm themselves with the knowledge and technology to distinguish between fake and authentic media since the distinctions between the two are getting increasingly hazy.

Thankfully, though, a number of creative approaches are emerging to make the voyage in the digital world easier, and these are covered here.

Deepfake Detection Tools:

Furthermore, while deepfake technology may be used lawfully for protected expression, education, and entertainment, relying just on detection tools is insufficient as a mitigating strategy. Nevertheless, it is important to consider the potential consequences and significance of these technical solutions. Early intervention and mitigation initiatives are facilitated by

[13] **Simran Jain & Piyush Jha,** *Deepfakes in India: Regulation and Privacy*, **South Asia@LSE** (May 21, 2020), https://blogs.lse.ac.uk/southasia/2020/05/21/deepfakes-in-india-regulation-and-privacy/.
[14] **Information Technology Act, 2000**, § 67, No. 21, Acts of Parliament, 2000 (India).

effective detection. Organisations in charge of transmitting multimedia content, such social media platforms, Internet Service Providers (ISPs), and other communication system providers, are in a prime position to evaluate the type of data they manage. There are precedents, like Microsoft's PhotoDNA technology, which detects copies of digital photos that have already been flagged and has been used by some providers to stop the spread of child sexual abuse materials (CSAM).[15] To strengthen team efforts against deepfakes, detection tool developments might be shared with social media businesses, ISPs, and communication system providers possibly as open-source resources.

Applying Deep-fakes to Promote Training and Authentication:

Deep-fake content creation and distribution can be a useful tool for increasing identification capabilities by training detection algorithms and human evaluators. On the other hand, there is a great chance to improve authentication methods in addition to initiatives to identify fake information. In particular, people and institutions can implement procedures to confirm and authenticate the veracity of the media they create and interact with. To promote the adoption of an open industry standard for content authenticity and provenance, for example, a group of media and technology companies, non-governmental organisations, academic institutions, and other stakeholders formed the Content Authenticity Initiative (CAI) in 2019. By allowing users to openly record the source and credit of their material, these standards make this information available to all those who follow them. As a result, users can check media for a "authenticity seal," increasing their trust in the information they come across. Commercial companies like Truepic also provide comparable authentication services, even though the CAI offers open standards that are available to all users. [16]

Harnessing Watermarking Technology for Digital Content Authentication:

Watermarking, a technique that involves directly embedding extra information or a host signal into the original content, is another way that technology can be used. This method creates secret keys inside the host data in order to identify alteration. The Amber Authenticate is a well-known cryptographic tool in this field that creates hashes at predetermined points during a

[15] **Fakhar Abbas & Araz Taeihagh,** *Unmasking Deepfakes: A Systematic Review of Deepfake Detection and Generation Techniques Using Artificial Intelligence*, 252 **Expert Syst. with Applications** 124260 (2024).
[16] **CHESA,** *Understanding C2PA: Enhancing Digital Content Provenance and Authenticity* (Sept. 2024), https://chesa.com/understanding-c2pa-enhancing-digital-content-provenance-and-authenticity/ (last visited June 14, 2025).

video, such a movie. As a result, any change made to the movie causes these hashes to be observably rearranged, alerting the audience to the change. Government organisations must proactively identify and remove officially recognised harmful materials because victims of synthetic media attacks, especially those involving non-consensual explicit content, find it difficult to remove content from all internet sources. In order to tag CSAM hash sets—a technique that could be used to combat dangerous deep-fakes governments have partnered with ISPs and social media companies.[17]

Conclusion & Recommendations:

It is critical to focus future research on creative solutions and strong defences as deep-fake technology develops further and poses more complex threats to digital authenticity and public trust. In order to lessen the negative effects of deep-fakes and promote a safer online environment, the following ideas highlight important areas of concentration, spanning from technological developments to teamwork and policy creation.

In order to guarantee that data is safe in an increasingly complicated digital environment, academics and stakeholders can collaborate to help them handle these crucial areas and keep ahead of new dangers.

- **Improvement of AI Detection Techniques:** More research must be done to create increasingly complex AI algorithms that increase the precision and efficacy of deepfake detection systems. First of all, it includes both the exploration of novel machine learning approaches and the advancement of existing ones, such as Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs).

- **Block-chain Integration for Verification:** Since block-chain technology has the potential to significantly improve authenticity verification processes, it is worthwhile to look into its ability to produce irreversible records of digital content.

- **Creation of Real-Time Detection Systems:** Research should concentrate on creating systems that can evaluate and validate digital content in real-time and give timely feedback on authenticity in order to counteract undesirable activities.

[17] **A. Qureshi, D. Megías & M. Kuribayashi,** *Detecting Deepfake Videos Using Digital Watermarking*, in **2021 Asia-Pacific Signal & Info. Processing Ass'n Annual Summit & Conf.** 1786 (Dec. 2021) (IEEE).

- **Encouragement of Collaborative Efforts:** Future studies should examine the efficacy and transparency of government and digital businesses working together to develop all-encompassing strategies to combat misinformation.

- **Investigation of User Empowerment Tools:** More research should be done on the development and distribution of easily accessible tools and educational resources meant to assist individuals in identifying deep-fakes and critically evaluating digital content.

- **Creation of AI Ethics Framework:** Thorough ethical frameworks must be built and continuously upheld in order to develop and apply generative AI technologies in an ethical manner. These frameworks should be founded on the fundamental principles of transparency, accountability for AI actions, and rigorous adherence to moral principles throughout the whole development and implementation process.

An analysis reveals an important omission: the creation of accurate, real-time digital copies came after the existing legal frameworks, such as common law torts, criminal statutes, and other legal frameworks. There are currently large loopholes in legal protection as a result of this technological innovation. In conclusion, the rapid advancement of deep-fake technology poses significant threats to individuals, businesses, and governments worldwide. Financial fraud, reputational damage, and the propagation of false information are all major risks associated with the misuse of deep-fakes. The concept of deep-fakes, the techniques used to produce them, and the potential benefits and drawbacks of this technology have all been covered in this Paper.

The legal system in India provides a few avenues for remedy, but it is also powerless to combat the deep-fake threat. The hazards posed by deep-fakes would decrease as a result of developing answers to these issues. In order to stop the propagation of false information, they should also emphasise media literacy and increase awareness of the potential danger posed by deep-fakes. establishing necessary technological protections, draughting precise and suitable rules and regulations, and educating people about the possible dangers and repercussions of deepfakes. Therefore, in order to ensure that the advantages of deepfake technology are realised while lowering the risks, the nations must work together to ensure that it is developed and deployed in a safe and responsible manner.