Combatting Deep-fakes in India – An Analysis of the Evolving Legal Paradigm and Its Challenges

Diya Sarkar[1]
Dr. Sudipta De Sarkar[2]

Abstract

Advancement of artificial intelligence and machine learning have spurred a new wave of propagation of false content and information about events and people, as part of entertainment, disparagement, fraud, influencing patterns and perceptions of consumers and voters etc., using deep-fake materials and targeted disinformation campaigns. Many institutions now perceive deep-fakes as a significantly greater hazard than identity theft, which can also be done with deep-fakes. This is especially true since the COVID-19 pandemic when most interactions went online. The advancement of deep-fake technology has reached a stage where the validity and integrity of any digital audio or video content available online may be called into question. This study presents a conceptualization of deep-fakes, explores their socio-legal ramifications, and evaluates the current legal ecosystem in the United States, Europe, and India. The authors through a comparative review intend to present constructive recommendations for addressing the difficulties posed by deep-fake technology and restoring trust within the digital ecosystem. The primary aim of the authors is to draw attention to the existing vulnerabilities linked to deep-fake technology and underscore the significance of implementing legislative regulations to effectively tackle these problems.

Keywords: *Artificial Intelligence, Generative Adversarial Networks, Regulatory initiatives.*

[1] Research Scholar, Kalinga Institute of Industrial Technology (A Deemed to be University), Bhubaneswar, Odisha, India.
[2] Kalinga Institute of Industrial Technology (A Deemed to be University), Bhubaneswar, Odisha, India.

I. Introduction

Deep-fakes pose a serious threat to all facets of society and the advent of artificial intelligence has boosted its malicious potential. With deep-fake technology, it has become exponentially easier to alter depiction of reality, with both the subject and the end consumer of the altered content falling unsuspecting victims to such practices. Audio and visual recordings of people saying and doing things they never said or did is at the core of such distortion. The proliferation of this technology is the result of the machine learning's increasing perfection, making it considerably harder to detect and prevent. While the technology has a number of advantages, it also has its share of drawbacks and can be misused at will. Since information interchange based on cognitive biases and algorithmic manipulation occurs frequently in the data driven economy, deep-fakes are a potent weapon of informational devastation faced by today's technology-driven civilization. It is anticipated that the spread of deep-fake technology will make the problem of truth decay in networked information exchange worse by each passing day.

From disparaging videos of movie actors[3] and sportsmen[4] and industrialists[5] in India, this menace will assume more sinister ambitions, and will be aimed at influencing the political climate and people's perception in the country. It is thus time that the open and daunting threat posed by unregulated deep-fake content is addressed head on. There are legal challenges towards regulating the production and distribution of synthetic video clips that imitate people without their express agreement. However, the lack of discernibility of such content from expressly created content, added by the speed with which it can be spread through the

[3] Ojha, A. (2024) 'Rashmika Mandanna Deep-fake: How Cops Traced Accused, Techie from Andhra's Guntur', *India Today*, 20 January. https://www.indiatoday.in/india/story/rashmika-mandanna-deep-fake-video-accused-arrested-andra-engineer-wanted-to-boost-followers-2491386-2024-01-20 (Last Accessed: 24 January, 2024).

[4] Livemint (2024) 'Sachin Tendulkar Becomes Latest Victim Of Deep-fake Video', 15 January. https://www.livemint.com/news/india/sachin-tendulkar-becomes-latest-victim-of-deep-fake-video-disturbing-to-see-11705308366864.html (Last Accessed: 17 January, 2024).

[5] Livemint (2023). *'False': Ratan Tata Calls Out A 'Deep-fake' Video Of Him Giving Investment Advice*. [online] mint. https://www.livemint.com/news/india/false-ratan-tata-calls-out-a-deep-fake-video-of-him-giving-investment-advice-11701926766285.html. (Last Accessed: 17 January, 2024).

internet, leaves no choice but to be controlled and regulated to prevent irreparable harm to both the subject and the end consumer of such content. Cybercrime, dissemination of false information, fraud, assault on privacy etc. are some of the potent areas where deep-fakes have been deployed both by individuals as well as organizations.

II. Emergence of Deep-fake

The concept of data manipulation is not recent. Misinformation via media dates a century back, to the 1890s when clunky 19th-century cameras made filming the Spanish- American War difficult for the Edison Manufacturing Company.[6] The production team mixed marching infantry and equipment with staged American soldiers destroying opposing units, inspiring patriotism amongst the American audience by the concealment of true facts behind the scenes.

The 1898 event was not a deep-fake, but it showed how data tampering could purposely disseminate altered information. Christoph Bregler, Michele Covell, and Malcolm Slaney pioneered Deep-fakes in 1997 with Video Rewrite[7] by means of a programme which edited videos to add words which no one featured in the video had uttered, and then automated the facial reanimation in sync with the added narrative. The event involving the video rewriting programmes led to the development of the Active Appearance Model (AAM) in 2001[8]. Additionally, the AAM computer vision programme compared a new image to a statistical model of the physiognomy and shape of an object. This study greatly improved facial matching, precision, and effectiveness[9].

---

[6] *The Spanish - American war in Motion Pictures* (n.d.) *The Library of Congress*. https://www.loc.gov/collections/spanish-american-war-in-motion-pictures/about-this-collection/ (Last Accessed: 12 January, 2024).

[7] www.historyofinformation.com. (n.d.). *Video Rewrite, Origins of Deep-fakes: History of Information*. [online] https://www.historyofinformation.com/detail.php?id=4792. (Last Accessed: 17 January, 2024).

[8] Cootes, T.F., Wheeler, G.V., Walker, K.N. and Taylor, C.J. (2002). View-based Active Appearance Models. *Image and Vision Computing*, [online] 20(9-10), pp.657–664. doi: https://doi.org/10.1016/s0262-8856(02)00055-0. (Last Accessed: 17 January, 2024).

[9] Cootes, T.F., Taylor, C.J., Cooper, D.H. and Graham, J. (1995). Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding*, [online] 61(1), pp.38–59. doi:https://doi.org/10.1006/cviu.1995.1004. (Last Accessed: 18 January, 2024).

In 2014, the Generative Adversarial Networks (GAN) was invented by Ian Goodfellow and a team of academics affiliated with the University of Montreal[10], and this was widely acclaimed for its utility across many applications, including enhancing astronomical photography within the scientific domain and facilitating the enhancement of video game quality by developers. But later on, the GAN made it possible for social media to post realistic video fakes, which have led to a surge of misinformation and diminished its benefits. Two Artificial Intelligence (AI) agents were included in GANs: one created a fake image, and the other found it. When the fake was detected, the forger AI grew by adapting and learning from its failures.

III. Understanding Deep-fake

Deep-fake is a combination of "deep learning" and "fake" that describes the use of artificial intelligence to create video content that is indistinguishable from reality to humans. Deep learning is an extensive AI technique that employs numerous layers of machine learning computations to extract progressively more complex features from unprocessed input.[11] It involves the utilization of algorithms that acquire knowledge from extensive datasets without the need for human supervision. As the size of the dataset increases, there is a higher likelihood that the algorithm will exhibit improved accuracy. It has the ability to learn from unstructured data, such as human facial expressions and features.[12]

For example, an AI can collect information about your physical movements[13] and by imitating the blinking patterns, head movements, speech patterns, and facial

[10] Kingra, S., Aggarwal, N. and Kaur, N. (2022). Emergence Of Deep-fakes And Video Tampering Detection Approaches: A Survey. *Multimedia Tools and Applications*. doi:https://doi.org/10.1007/s11042-022-13100-x. (Last Accessed: 18 January, 2024).

[11] Schmidhuber, J. (2015). Deep Learning In Neural Networks: An Overview. *Neural Networks*, 61(61), pp.85–117. doi:https://doi.org/10.1016/j.neunet.2014.09.003. (Last Accessed: 15 January, 2024).

[12] Heo, Y.-J., Yeo, W.-H. and Kim, B.-G. (2022). Deep-fake Detection Algorithm Based On Improved Vision Transformer. *Applied Intelligence*. doi:https://doi.org/10.1007/s10489-022-03867-9. (Last Accessed: 15 January, 2024).

[13] Waseem, S., S. A. R. Abu-Bakar, Omar, Z., Bilal Ashfaq Ahmed, Baloch, S. and Adel Hafeezallah (2023). Multi-Attention-Based Approach For Deep-fake Face And Expression Swap Detection And Localization. *Eurasip Journal on Image and Video*

expressions, individuals appearing in videos are used to create duplicates that are uncannily identical but fake.[14]

The basic essence of a deep-fake is false information, manipulation, or fabrication, because deep-fakes seek to manipulate our perception of reality and convince us that something is true.[15] Any piece of media (video, audio, or otherwise) that has been doctored or constructed from scratch (in whole or in part) is considered a deep-fake. It is possible to employ a variety of technologies for this, but the most commonly used is based on GAN[16]. According to a Dutch report[17], the GAN technology has evolved in terms of content quality and resolution to the point that the image, audio or video processed through it won't be an exact reproduction but will look authentic. A video that pretends to show someone saying or doing things they have never done or mentioned could be made in a matter of minutes. Such footage could be difficult to discern from authentic stuff.[18]

IV. Mapping Deep-fake and its Implications

Late in 2018, the populace residing in Gabon, a nation situated in central Africa, had been deprived of any public appearances of their President Ali Bongo for an extended period of time. There then existed conjecture on the potential concealment by the government of his illness or demise. The government

Processing, 2023(1). Doi : https://doi.org/10.1186/s13640-023-00614-z. (Last Accessed: 15 January, 2024).

[14] Engler, A. (2019). *Fighting Deep-fakes When Detection Fails*. [online] Brookings. Available at: https://www.brookings.edu/articles/fighting-deep-fakes-when-detection-fails/. (Last Accessed: 20 January, 2024).

[15] Appel, M. and Prietzel, F. (2022). The Detection Of Political Deep-fakes. *Journal of Computer-Mediated Communication*, 27(4). doi:https://doi.org/10.1093/jcmc/zmac008. (Last Accessed: 15 January, 2024).

[16] *Supra Note 9.*

[17] Van Der Sloot, B., Wagensveld, Y., and Koops, B.-J. (2021). *Deep-fakes: The Legal Challenges Of A Synthetic Society*. [online] Available at: https://www.tilburguniversity.edu/sites/default/files/download/Deep-fake%20EN.pdf. (Last Accessed: 15 January, 2024).

[18] Department of Homeland Security (2023). *Increasing Threat of Deep-fake Identities*. [online] Available at: https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deep-fake_identities_0.pdf. (Last Accessed: 16 January, 2024).

officially quelled the conjectures by issuing a statement confirming that Bongo had suffered a cerebrovascular accident (stroke) but otherwise was healthy. Subsequently, the government disseminated a video recording of his customary annual address on the New Year's Eve.[19] The forensic investigation of the video did not uncover any alterations or manipulations. That had no relevance. The mere concept of deep-fakes was sufficient to exacerbate an already pre-carious situation.

Deep-fake attacks also include social engineering, such as the so-called CEO fraud attack (also known as business email compromise), in which an attacker poses as a government official or a superior and requests a money transfer. In 2019, cybercriminals employed AI based software to emulate the voice of a chief executive, orchestrating a fraudulent transfer of €220,000 in Europe. Factually it was the Chief Executive Officer (CEO) of an energy company headquartered in the United Kingdom (U.K.) who mistakenly believed he was engaged in a telephone conversation with his superior, the CEO of the company's German parent organization. During this conversation, the UK CEO believed that the German CEO requested him to transfer funds to a supplier located in Hungary.[20] This incident, as noted by experts in the field of cybercrime, represents a distinctive occurrence, wherein deep-fake voice-spoofing enabled by AI technology was leveraged for hacking purposes. Conventional cybersecurity solutions, which are primarily geared to safeguard corporate networks against unauthorized access, were unable to detect manipulated or counterfeit voices.

As deep-fake technology becomes more convincing, fears of political impact from fake media had grown ahead of the 2020 US presidential elections.[21] A

[19] Cahlan, S. (2020). *How Misinformation Helped Spark An Attempted Coup In Gabon. [online] Washington Post*. Available at: https://www.washingtonpost.com/politics/2020/02/13/how-sick-president-suspect-video-helped-sparked-an-attempted-coup-gabon/. (Last Accessed: 16 January, 2024).

[20] Damiani, J. (2019). *A Voice Deep-fake Was Used To Scam A CEO Out Of $243,000*. [online] Forbes. Available at: https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deep-fake-was-used-to-scam-a-ceo-out-of-243000/?sh=2840c2c92241. (Last Accessed: 16 January, 2024).

[21] Villasenor, J. (2019). *Deep-fakes, Social Media, And The 2020 Election*. [online] Brookings.edu. Available at: https://www.brookings.edu/articles/deep-fakes-social-media-and-the-2020-election/. (Last Accessed: 16 January, 2024).

recent analysis from Deeptrace Labs, a cybersecurity startup that detects deceit, identified no known disinformation efforts using deep-fakes.[22] The biggest influence has been realising they can be used in that manner. Deep-fakes threaten politics by making fake media appear real, but the bigger worry is how they might be used to make the real look fake.[23] Perhaps, the utilization of bots by the attackers to respond to the inquiries posed by the victim remains ambiguous. In the event that such actions were undertaken, it is plausible that the task of law enforcement authorities in conducting investigations have been further complicated.

Deep-fakes has a wide range of potential applications, encompassing the substitution of a different visage onto pre-existing video content featuring another individual's facial attributes. Such deep-fake incidents can be traced through a deep-fake video circulated in 2022 on Twitter, purportedly depicting Russian President Vladimir Putin making a declaration of peace.[24]

In recent developments, it has come to light that both Meta and YouTube have made the decision to remove a deep-fake video featuring Ukraine's President discussing the possibility of surrendering to Russia.[25] The lack of persuasiveness in President Zelensky's fake temperament was received with disdain by a significant number of Ukrainian citizens. However, the Ukrainian Centre for Strategic Communications then issued a warning over the potential utilization of deep-fakes by the Russian government as a means to persuade Ukrainians to

[22] Adjer, H., Patriani, G., Cavalli, F. and Cullen, L. (2019). *The State Of Deep-fakes Landscape, Threats, And Impact*. [online] Regmedia.co.uk. Available at: https://regmedia.co.uk/2019/10/08/deep-fake_report.pdf (Last Accessed 12 January, 2024).

[23] World Economic Forum. (2020). *Deep-fake Democracy: Here's How Modern Elections Could Be Decided By Fake News*. [online] Available at: https://www.weforum.org/agenda/2020/10/deep-fake-democracy-could-modern-elections-fall-prey-to-fiction/. (Last Accessed: 20 January, 2024).

[24] Butler, S. (2022). *Putin Declares Peace With Ukraine In Debunked Deep-fake Video*. [online] indy100.com. https://www.indy100.com/news/putin-peace-ukraine-deep-fake-video. (Last Accessed: 16 January, 2024).

[25] Holroyd, M. (2022). *Deep-fake Video Shows Zelenskyy's False Call For Ukraine To Surrender*. [online] euronews. https://www.euronews.com/my-europe/2022/03/16/deep-fake-zelenskyy-surrender-video-is-the-first-intentionally-used-in-ukraine-war. (Last Accessed: 20 January, 2024).

capitulate. Given that both sides have been using modified media content, we are free to draw conclusions from these films about the amount of false information that is present in the conflict.

In addition to the generation of highly accurate deep-fake audio that faithfully replicates the voice of a certain person, the phenomenon was observed in two emulations of the renowned podcaster Joe Rogan's voice. The first, as reported in a 2019 emulation, was generated by the utilization of a deep learning system that converted written text into synthesized speech.[26] The synthesized "voice" of Rogan was employed to discuss his purported sponsorship of a hockey team comprised entirely of chimpanzees. And the other was in 2023, where in an advertisement, Rogan engages in a conversation with another person regarding the topic of male supplements, specifically focusing on a brand called "Alpha Grind". Rogan purportedly endorsed the efficacy of the drug, highlighting its prominent ranking on Amazon as one of the highest-rated items.[27] It's crucial to remember that Rogan said nothing of the like on his podcast and the commercial used an AI rendition of Rogan's voice endorsing the goods without his knowledge or express approval.

V. Potential Risks and Severity of Deep-fakes

Deep-fake scams have emerged as a contemporary security menace targeting both individuals and corporations. It is utmost necessary to implement appropriate measures to safeguard and fortify the individuals or institutions against potential deep-fake-related threat.[28] Due to the prevalence of individuals

[26] Thompson, S.A. (2023). *Making Deep-fakes Gets Cheaper and Easier Thanks to A.I*. [online] 12 Mar. https://www.nytimes.com/2023/03/12/technology/deep-fakes-cheapfakes-videos-ai.html#:~:text=Meme%2Dmakers%20and%20misinformation%20peddlers. (Last Accessed: 16 January, 2024).

[27] Jay, S. (2023). *Rap, Reddit, Rogan, and Repercussions – Navigating Deep-fake Legality in a Viral World.* [online] sports-entertainment.brooklaw.edu. https://sports-entertainment.brooklaw.edu/media/rap-reddit-rogan-and-repercussions-navigating-deep-fake-legality-in-a-viral-world/ (Last Accessed: 16 January, 2024).

[28] Federal Office for Information Security. (n.d.). *Deep Fakes – Threats and Countermeasures*. [online]: https://www.bsi.bund.de/EN/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deep-fakes/deep-fakes_node.html. (Last Accessed: 18 January, 2024).

sharing numerous videos and audio recordings of themselves on social media platforms, scammers are able to exploit easily accessible techniques to manipulate and deceive them into believing that they are engaging in actions and uttering statements that they have not actually performed. There exist two principal dangers that are related with this occurrence – National and International, and, Societal.

A. National and International Ramifications

Deep-fake technologies possess the potential to be deceitful, malicious, and even catastrophic at various levels, including the individual, organizational, and social domains. As to the findings of the Brookings Institute, the advent of deep-fakes presents a substantial menace to both the integrity of democratic discourse and the gradual erosion of public trust in institutions.[29]

The proliferation of deep-fakes poses a significant challenge to global as well as national stability, encompassing the dissemination of misinformation and substantial cybersecurity vulnerabilities. As discussed, deep-fakes pose a significant concern for the upcoming elections, giving rise to substantial inquiries regarding the credibility of democratic elections, policy formulation, and the broader fabric of our society. It is plausible that state opponents or individuals driven by political motivations may disseminate manipulated recordings depicting elected officials or other prominent figures engaging in inflammatory discourse or exhibiting inappropriate conduct. Engaging in such actions has the potential to undermine the confidence of the public, have adverse consequences on the quality of public opinions, and potentially influence the outcome of an electoral process.

The intelligence community of U.S.A found that Russia, for the purpose of undermining public confidence in the democratic process, conducted significant "influence operations" during the 2016 US Presidential elections[30]. These

[29] Galston, W.A. (2020). *Is Seeing Still Believing? The Deep-fake Challenge To Truth In Politics*. [online] Brookings. https://www.brookings.edu/articles/is-seeing-still-believing-the-deep-fake-challenge-to-truth-in-politics/ (Last Accessed: 19 January, 2024).

[30] Office of The Director of National Intelligence (2017). *Background to Assessing Russian Activities and Intentions in Recent US Elections: The Analytic Process and Cyber Incident Attribution*. [online]

activities sought to discredit the then US Secretary of State, Ms. Hillary Clinton, affecting her Presidential prospects. The Russian government was alleged to have employed state-funded media platforms and paid internet forums to undermine the electoral prospects of the Clinton presidential campaign. In addition to its implications within domestic politics, the utilization of deep-fake technologies presents a significant apprehension for the diplomatic efforts and national security nations at the global front. The preceding discussion over the video circulated on social media in 2022, featuring Ukrainian President Zelensky purportedly issuing directives to his troops to surrender to Russian military forces, has been classified as a deep-fake.[31]

According to specialists, although the current deep-fake under consideration did not possess a high level of sophistication, there was a possibility that forthcoming advancements in audio or video forgery techniques could enhance the effectiveness of malevolent influence operations.

B. Societal Ramifications

False information spread via digital avatars sharpens social divisions. If majority of digital content is manipulated, the media will need a lot of time and resources to authenticate it, identify the specific manipulations in videos, images, and other media, and evaluate their impact on the news item. An increasing amount of information eluding detection can have serious detrimental consequences on the below mentioned two areas-

i. **False reporting and diminishing public trust –**

Deep-fakes spread misinformation, escalating social tensions. The immediate concern is about the growing volume of undetected data, within the sphere of what has been labelled as “fake news”. The mass media can acknowledge a margin of error, leading to legitimate accusations of “fake news” or follow strict guidelines and protocols, falling behind media organizations that quickly release sensational (and potentially unreliable) news pieces for public consumption.

---

https://www.dni.gov/files/documents/ICA_2017_01.pdf. (Last Accessed: 19 January, 2024).

[31] *Supra Note 24*.

ii. **Impact on Women's Safety and Social Status**

Deep-fake content can be a major contributor to the exacerbation of sexism and misogyny, perpetuating the stigmatization of women in society. Deep-fake content has the potential to subject the victim, in most cases women, into becoming the targets of sexual objectification, with deep-fakes being utilized to disseminate non-consensual explicit content of such women. Hence, the prevalence of misogynistic remarks and the act of public shaming is already extensive in both offline and online contexts, and the advent of unchecked deep-fake technology will inevitably aggravate these issues. The need for private recordings has diminished due to the ease with which individuals can fabricate and distribute pornographic content featuring others as a means of seeking retribution, either among acquaintances or within online social networks. Scholars emphasize that the knowledge of some content being misleading holds relatively limited significance. Despite the awareness among individuals that the explicit content in question is a manipulated digital representation, the potential consequences on society can be profoundly detrimental. These activities are outright damning to an individual's sense of self and self-esteem, and any post facto redressal will not be able to compensate wholly for such loss.

VI. Looming Legal Issues of Deep-fakes and the Regulatory Standards

Deep-fakes give rise to a range of psychological, financial, and sociological challenges. Presently, there exists no established legal framework within criminal law or civil responsibility in India that can outright tackle the problems of production or dissemination of deep fake content. The desirability of implementing a complete prohibition on digital manipulation is questionable, as the inherent problematic nature of digital manipulation is not universally acknowledged.

In a circuitous route, the World Intellectual Property Organization (WIPO) in 2019 released a draft on Intellectual Property Policy and Artificial Intelligence which addressed two specific questions pertaining to the subject of deep-fakes. According to the WIPO, deep-fakes have the potential to give rise to more significant issues, such as the infringement of human rights, right to privacy, and the protection of personal data, as compared to copyright violations. WIPO

observed that in cases when deep-fake content is shown to be entirely inconsistent with the individual's actual experiences and circumstances, it was deemed inappropriate to grant copyright protection to such work. Accordingly, in the context of copyright, it suggested that ownership of deep-fakes should be attributed to their creator. Deep-fakes are produced without source person's input in terms of image, sound, or other properties due to consent requirements. Thus, copyright was ineffective in combating deep-fake technology since the subject victims lacked copy-right ownership.[32]

However, in the present scenario, attribution of copyright in case of deep-fakes will be of lesser priority as compared to the damage that can be caused by such to an individual or society, for which a dedicated legal framework involving criminal and civil penalties has to be considered on high priority. In light of this, it is necessary to peruse the laws and legal frameworks available in jurisdictions like the U.S.A and the E.U, which can be of guidance for India when she decides to handle the issue from a legal standpoint.

A. Deep-fake Regulations in United States

The United States of America (U.S.A.) took the initiative as the pioneering nation in addressing the emergence of artificial intelligence technologies. The Malicious Deep Fake Prohibition Act was enacted by the United States Congress in 2018. This legislation holds significance as it marks the initial attempt to provide a legal definition for the term “Deep-fake”. The DEEP-FAKES Accountability Act was officially introduced in 2019. Nevertheless, the public raised concerns and objections regarding its imprecise delineations and its potential infringement upon the First Amendment of the United States Constitution.

During that year itself, the Congress put out a proposition referred to as the Deep-Fake Report Act of 2019. This proposal mandated that the United States Department of Homeland Security should be obligated to periodically produce assessment reports pertaining to the technology known as deep-fakes. Currently,

[32] WIPO Secretariat (2019). *WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI) Second Session: Draft Issues Paper On Intellectual Property Policy and Artificial Intelligence*. [online] www.wipo.int.https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_2_ge_20/wipo_ip_ai_2_ge_20_1.pdf (Last Accessed: 14 January, 2024)

there exists a lack of comprehensive federal legislation that adequately tackles the potential concerns associated with deep-fake technology. In December 2019, Section 5709 of the National Defense Authorization Act (FY23) obligated the Director of National Intelligence to provide a report on the use of deep-fakes by foreign governments, their ability to disseminate misinformation, and their possible implications upon national security. While this law effectively governs the issue of deep-fakes originating from foreign entities within the United States, it fails to adequately tackle the internal deep-fake predicament within the country.

The National Defence Authorization Act (NDAA) holds the distinction of being the first legislation to be enacted, encompassing provisions pertaining to deep-fakes. However, subsequent to its passage, two other bills have been approved by one Congressional chamber and are currently awaiting consideration by the other. One of the legislative proposals under consideration is the Identifying Outputs of Generative Adversarial Networks (IOGAN) Act, officially known as H.R. 4355. It requires the National Science Foundation (NSF) and the National Institute of Standards and Technology (NIST) to allocate funds and resources towards doing research on modified or synthesized media, which includes the results produced by GAN.[33]

The Senate has successfully passed the other bill i.e., Deep-fake Report Act of 2019, also known as Bill S. 2065. However, its progress is now dependent on its passage in the House of Representatives. It pertains to the issuance of a report by the Department of Homeland Security about digital content forgery technologies.[34] According to Matthew Ferraro, in recent years, thirteen states have enacted legislation specifically addressing this type of information, resulting in a complex combination of civil and criminal measures that pose challenges for enforcement. Ferraro asserts that there have been no recorded instances of legal charges being filed against any American person for the

[33] Titles - H.R.133 - 116th Congress (2019-2020): Consolidated Appropriations Act, 2021, H.R.133, 116th Cong. (2020), https://www.congress.gov/bill/116th-congress/house-bill/133/titles. (Last Accessed: 22 January, 2024).
[34] S.2065 - 116th Congress (2019-2020): Deep-fake Report Act of 2019, S.2065, 116th Cong. (2019), https://www.congress.gov/bill/116th-congress/senate-bill/2065. (Last Accessed: 22 January, 2024).

creation of AI-generated non-consensual sexualized content.[35] Therefore, as observed, certain state jurisdictions exhibit prompt responses to instances of the inappropriate utilization of deep-fake technology, particularly in relation to "pornographic videos" and "political elections".[36]

In 2019, the state of Texas became the first jurisdiction in US to enact legislation that prohibited the creation and distribution of deep-fake videos, specifically with the intention of inflicting harm onto candidates running for public office or influencing the electoral outcomes. Within 30 days of an election, it is now a Class A misdemeanour in Texas to create and distribute a deep-fake video with the aim to harm a candidate or manipulate the election result. Imprisonment for one year in county jail and a $4,000 fine is the maximum punishment for this violation.

Subsequently in 2019, California enacted two laws which granted victims of non-consensual deep-fake pornography the right to pursue legal action for compensatory damages. Additionally, they provided candidates running for public office with the opportunity to file lawsuits against individuals or organizations that disseminate election-related deep-fakes without appropriate warning labels, specifically in cases where such actions are carried out with deliberate intent to cause harm, commonly referred to as "actual malice", in close proximity to Election Day.

The Calif. AB-1280 bill [Crimes: deceptive recordings] failed passage. In the same year, legislations were enacted by Georgia and Virginia to criminalize the production and distribution of non-consensual deep-fake pornography. In 2020, New York enacted legislation that established legal provisions pertaining to the pursuit of legal remedies against the unauthorized dissemination of deep-fake content. The introduction and implementation of AI regulation measures in

[35] D'Anastasio, C. and Alba, D. (2023). *Google and Microsoft Are Supercharging AI Deep-fake Porn*. [online] bloomberg.com. Available at: https://www.bloomberg.com/news/articles/2023-08-24/google-microsoft-tools-behind-surge-in-deep-fake-ai-porn. (Last Accessed: 23 January, 2024).

[36] Ferraro, M. (2019*). Deep-fake Legislation: A Nationwide Survey—State and Federal Lawmakers Consider Legislation to Regulate Manipulated Media*. [online] WilmerHale. https://www.wilmerhale.com/insights/client-alerts/20190925-deep-fake-legislation-a-nationwide-survey. (Last Accessed: 23 January, 2024).

various jurisdictions had increased in 2022.[37] However, majority of the states lack comprehensive AI legislation, with only the five states named above having deep-fake legislation. As state and federal governments cope with emerging technology that potentially affect persons, companies, and society, laws in this field are changing swiftly.[38]

Sl. no.	Legislation type	Regulation	Year	Deep-fake Related Content
1.	**Federal**	Malicious Deep-fake Prohibition Act	2018	Set up criminal office relating to deep-fake election media.
2.	**Federal**	Deep-fakes Accountability Act	2019	Decides civil & criminal accountabilities for altered media.
3.	**Federal**	The Deep-fake Report Act	2019	Mandates reporting annually on deep-fake pornography.
4.	**Federal**	National Defense Authorization Act (NDAA)	2019	Imposes a reporting requirement and notification provision.
5.	**State- Texas**	Tex. SB 751	2019	Deep-fakes for causing harm to electoral process or public officials.

[37] Poritz, I. (2023). *States Are Rushing to Regulate Deep-fakes as AI Goes Mainstream*. Bloomberg.com. [online] 20 Jun. https://www.bloomberg.com/news/articles/2023-06-20/deep-fake-porn-political-ads-push-states-to-curb-rampant-ai-use. (Last Accessed: 24 January, 2024).

[38] Lipkowitz, D. (2020). *Manipulated Reality, Menaced Democracy: An Assessment Of The Deep Fakes Accountability Act Of 2019* – N.Y.U. Journal of Legislation & Public Policy. [online] https://nyujlpp.org/quorum/lipkowitz-manipulated-reality-menaced-democracy-deep-fakes-accountability-act/ (Last Accessed: 24 January, 2024).

6.	**State-California**	a. Calif. AB-602 [Individual depicted using digital or electronic Technology: Sexually Explicit Content] b. Calif. AB-730 [Misleading Audio/Video During Elections] c. CA AB2658 [Assembly Bill]	a. 2019 b. 2019 c. 2022	a. Nonconsensual sexually explicit (porn) Deep-fake. b. Deep-fakes related to elections. c. Protection to juveniles on electronic monitoring devices.
7.	**State-Georgia**	(2019 GA. S.B. 337)	2019	Bans the online dissemination
8.	**State-Virginia**	Code of Virginia-Unlawful Dissemination or Sale of Images of Another Person	2019	Criminal sanctions for non-consensual deep-fake pornography.

9.	**State- New York**	N.Y. A08155, S0587 –B	2020	Commercial exploitation, nonconsensual pornography, and commercial exploitation. It created a right of publicity barring the commercial use of a deceased performer's digital replica until 40 years had passed since his death.
10	**State- Washington**	SB 6280 Act	2020	Safeguards for use of facial recognition
11	**Federal Bill**	IOGAN Act (H.R. 4355)	2019	to support research and development relating to deep-fakes.
12	**State- Massachusetts**	An Act to Protect Against Deep Fakes Used to Facilitate Criminal or Tortious Conduct Bill H. 1755	2021	Expand identity fraud definition to criminalize the creation or distribution of deep fakes with malicious or wrongful intent.
13	**Federal Bill**	Deep-fake Task Force Act	2021	To address and reduce threats posed by digital content forgeries.

Source – Compilation of Data from US Statutes

B. Deep-fake Regulations in Europe

The European Union (EU) has adopted a proactive stance towards the regulation of deep-fakes, advocating for heightened research efforts in the areas of deep-fake identification and prevention. The recent Artificial Intelligence Act, 2024

(AIA) in the EU has taken a very proactive stance towards combatting deep-fakes. In addition to defining deep-fakes as "manipulated or synthetic audio, image or video content that would falsely appear to be authentic or truthful, and which features depictions of persons appearing to say or do things they did not say or do, produced using AI techniques, including machine learning and deep learning" (Proposed Amendment 44d to Article 3 para 1), the Act also imposes obligations on the creator of deep-fake or the user of the AI system, "*generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful" that the "the content has been artificially generated or manipulated*" (Article 52(3)). Critics however state that mere disclosure of the nature of the deep-fake will not be of any avail when it is intended to degrade the esteem and persona of its subjects, especially women. Furthermore, it is possible to circumvent the regulations, and no special obligations have been imposed on online and digital forums responsible for creation and dissemination of such content.[39]

The EU AIA tackles the problem of deep-fake by incorporating three main components:

(a) A clear definition of deep-fake (Article 3(60));

(b) Requirements for AI providers and deployers, who are also known as deep-fake creators, to be transparent (Article 50); and,

(c) Recitals 132 to 137.

Article 3(60) of the document provides a definition for "deep-fake" as any artificially created or altered image, audio, or video content that is designed to appear genuine or authentic, and closely resembles real individuals, locations, objects, events, or entities. The AIA imposes more stringent regulations for higher levels of risk.

Additionally, the EU had earlier proposed regulatory measures mandating the transparent labelling of artificially generated content to address deep-fakes,

[39] Łabuz, M. (2023). Regulating Deep Fakes in the Artificial Intelligence Act. *Applied Cybersecurity & Internet Governance*, 2(1). doi:https://doi.org/10.60097/acig/162856. (Last Accessed: 25 January, 2024).

which has been duly recognized under the newly enacted AIA. The 133 recital of the AIA acknowledges the necessity of adaptability to handle diverse content formats, advanced technologies, and artificial intelligence capabilities. This guarantees effective adherence for service providers, particularly those handling varied material and advancing technology. Recital 133 highlights the significance of precise, compatible, and efficient techniques for labelling and identifying content. Their purpose is to monitor the source of content and provide evidence of its authenticity. In accordance with Article 50(4), it is mandatory for producers, artists, or anyone else to disclose the use of AI in the creation of deep-fakes. Such disclosure should be easily discernible through distinct labels and references to the AI source, as Recital 134 stresses, in order to ensure transparency. The two exceptions to this rule, as stated in Article 50(4), are

(a) cases involving the investigation of crimes or the collection of evidence, and,

(b) cases involving the protection of artistic production and freedom of speech as outlined in Articles 11 and 13 of the EU Charter.

A "limited disclosure obligation" that guarantees transparency without compromising artistic expression is mentioned in Recital 134 with regard to deeply faked works of art or movies that are demonstrably creative, fictional, satirical, or artistic. However, Labuz in his work has highlighted that exemptions for humor in freedom of speech could potentially lead to a legal ambiguity where destructive narratives are presented as satire.[40]

The European Commission in 2018 had released a detailed report titled "Tackling online disinformation: a European Approach" which presented a set of regulations aimed at curbing the unlawful manipulation of public opinion by information providers.[41] In 2021, the panel for the future of science and technology studied and submitted a report titled "Tackling Deep-fakes in European policy", wherein it outlined distinct elements of the deep-fake lifecycle

[40] *Ibid.*

[41] European Commission (2018) Tackling online disinformation: a European Approach. *Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions* [online] Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0236. (Last Accessed: 24 January, 2024).

that could be considered by policy-makers in order to mitigate and respond to the negative consequences associated with deep-fakes. The study further discussed several gaps in the regulatory landscape related to deep-fake. While discussing the exemptions under then proposed AI regulation, the study indicated a high risk in terms of deep-fake detection software use by law enforcement authorities as it had the potential to pose a threat to the rights and freedoms of the individuals.

Based on research, the objective was to establish an equilibrium that fostered transparency and accountability while safeguarding creative autonomy and ingenuity. Although it does not explicitly mandate overall content labelling, The AIA encourages openness and accountability through many means that can indirectly impact the presentation and perception of AI-generated material.[42]

There are stringent limitations on deep-fake as a result of the EU's General Data Protection Regulation (GDPR). According to Article 4(19) of the law, personal data is defined as "any information that pertains to a specific or identifiable individual". If a deep-fake represents a real person, it is unquestionably within the scope of the GDPR because it can be easily recognized as that individual.[43] An individual whose personal information has been compromised by deep-fake technology can therefore pursue legal recourse under GDPR. The GDPR requires that individuals' records be accurate and, when applicable, updated on a timely basis (Article 5 (1)). It is the responsibility of data controllers to take appropriate measures to ensure that any personal data found to be inaccurate, considering the purposes for which they are processed, are promptly rectified, or erased.

Even if the developers aren't building deep-fakes directly, they could still face legal consequences if they exploit personally identifiable information to train algorithms. In the same vein, producers and sellers can be held accountable if they utilize this information to build and sell deep-fakes. It boils down to this:

---

[42] Ghaswalla, S. (2023). *Can Content Labels Combat the Threat of Malicious Deepfakes and Fake News?* [online] Medium. https://sorabg.medium.com/can-content-labels-combat-the-threat-of-malicious-deepfakes-and-fake-news-726b88bd05be (Last Accessed: 22 January, 2024).

[43] Felipe Romero Moreno (2024). Generative AI And Deepfakes: A Human Rights Approach To Tackling Harmful Content. *International Review Of Law Computers & Technology*, pp.1–30. doi: https://doi.org/10.1080/13600869.2024.2324540. (Last Accessed: 23 January, 2024).

the GDPR is not concerned about the various deep-fake development stages and strict compliance is necessary only when dealing with personal data.[44] Under Article 6, GDPR, the justification for consent and legitimate interests for creating deep-fakes is relevant. The CJEU states that in order for consent to be valid, it must be explicit, unambiguous, freely offered, and informed. Therefore, when attempting to use personal data for deep-fakes, it is necessary to secure the individual's consent.[45] Furthermore, if the deep-fake content is indeed authentic and reliable, an individual whose personal data has been manipulated in a deep-fake scenario, commonly referred to as a "data subject", has the potential to exercise the right to erasure as outlined in Article 17, GDPR if the deep-fake information is genuinely authentic and reliable.

In addition to the existing range of regulations and initiatives aimed at integrating deep-fake within the existing regulatory framework, EU has come up with the world's first comprehensive AI law and it seems to be well poised in its efforts at combatting the rising threat from deep-fakes. These measures serve to restrict the use of deep-fake technology in the context of disinformation governance, safeguarding individual information, and regulating artificial intelligence. The European Commission's AI regulatory framework is the most significant for law enforcement in deep-fakes, and its application is awaited.

The EU has also enacted legislative measures that mandate social media corporations to eliminate deep-fakes and other forms of disinformation from their platforms. It's Code of Practice on Disinformation, which was revised in June 2022 vis., The Digital Services Act 2022 (DSA), includes provisions aimed at addressing the issue of deep-fakes. Since February 2024, the DSA 2022 became applicable to all EU member States affecting European users who create and disseminate online content as well as tech companies who act as intermediaries on the internet. Violators of these provisions may face fines of up to 6 % of their global income. The code was initially presented in 2018 as a voluntary self-

[44] European Parliament (2021). *Tackling Deepfakes In European Policy*. [online] https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf. (Last Accessed: 23 January, 2024).

[45] Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband e.V. v Planet49 *GmbH* [2019] 22 C-673/17. (Last Accessed: 16 January, 2024).

regulatory document, but it has gained support from the DSA. The DSA, enhances the surveillance of digital platforms to address several forms of misuse.[46]

The legislative framework pertaining to AI put forth by the European Commission offered a potential avenue for addressing some dangers. However, it is important to note that regulations should not solely concentrate on the technological aspects of deep-fakes. The study encompasses many policy choices within each of the dimensions. The analysis shows that the EU's deep-fake regulatory environment is marked by a sophisticated network of constitutional principles and strict and flexible rules at both the supranational EU and Member State levels. These options have the potential to be integrated into the AI legislative framework, as well as the planned European Union digital services act package and other relevant contexts. It is probable that a comprehensive set of steps will be required to mitigate the hazards associated with deep-fakes, while simultaneously capitalizing on their potential[47].

The EU Code of Practice on Disinformation, which was endorsed by the European Council in 2018, advocates for self-regulation within the sector, imposes limitations and oversight on the use of illicit "deep-fake" synthesis technology, and safeguards the personal data, including photos, of individuals. The most recent endeavours originate from the preeminent technical standards agency in Europe, which has published a report titled ETSI GR SAI 011, discussing the potential hazards associated with employing artificial intelligence for the purpose of manipulating digital identity representations[48]. ETSI has added that deep-fake attacks are targeting remote biometric identification and authentication. Video-based remote identification is a prevalent practice

[46] Paemen, D. (2022). *Digital Services Regulation In The EU: An Evolving Landscape*. [online] https://www.cliffordchance.com/briefings/2022/09/-digital-services-regulation-in-the-eu--an-evolving-landscape.html (Last Accessed: 22 January, 2024).
[47] *Supra Note 40*.
[48] Securing Artificial Intelligence (SAI); *Automated Manipulation of Multimedia Identity Representations Disclaimer*. (2023). https://www.etsi.org/deliver/etsi_gr/SAI/001_099/011/01.01.01_60/gr_SAI011v010101p.pdf. (Last Accessed: 23 January, 2024).

employed by numerous European financial institutions to facilitate the account opening process for consumers and ensure adherence to regulatory requirements.

In addition to their primary function, speaker recognition systems are utilized for the purpose of verifying the identity of consumers who are seeking to initiate transactions. It warns there is a large variation in the security quality of these procedures and their susceptibility to assaults involving modified identity representations. Therefore, Attacks may encompass the unauthorized acquisition of biometric data from individuals without their awareness, or alternatively, may rely on entirely fabricated synthetic data. It reported that deep-fakes pose a complex problem, for which there is no silver bullet, but which can be combated most effectively through a combination of measures on multiple levels. ETSI presents several strategies to address the issue of deep-fake technology, encompassing educational initiatives, awareness campaigns, and the implementation of regulatory measures mandating the labelling of modified identity depictions.[49]

Table 2. European Union Deep-fake Regulations
(Apart from the provisions in the Artificial Intelligence Act, 2021)

Sl. no.	Regulation / Statute	Year	Deep Fake Related Content
1.	The Artificial Intelligence Act	2024	A comprehensive regulatory framework for AI that encompasses the issue of deep-fake technology.
2.	The Digital Services Act	2022	It regulates online intermediaries and platforms.
3.	The AI Regulatory Framework	2021	Labelling deep-fake content to warn users of modified footage.
4.	General Data Protection Regulation	2018	Data protection and privacy.
5.	Tackling online disinformation: a	2018	Combating online misinformation which

[49] Dahmen-Lhuissier, S. (n.d.). *ETSI - Best Security Standards | ETSI Security Standards*. [online] ETSI. https://www.etsi.org/technologies/securing-artificial-intelligence. (Last Accessed: 16 January, 2024).

	European approach		manipulates public opinion
6.	Code of Practice or Action Plan on Disinformation	2018	Incentivize web platforms to monitor deep-fake and enforce penalties upon violators.
7.	e-Commerce Directive	2000	Addresses challenges pertaining to disinformation, misinformation and the proliferation of fake news.
8.	Digital Services Act	2022	Surveill digital platforms to handle misuse and eliminate fake content.
9.	Audio Visual Media Directive	2020	Addressing animosity and safeguarding cultural pluralism.
10.	Copyright regime	2021	Addresses fair use
11.	Democracy Action Plan	2020	Increase media freedom, support fair elections, and fight disinformation.

Source – Compilation of data from EU Statutes / Regulations

C. Lessons for India

The Government at the moment is waking up to the threat posed by deep-fakes and its implications across a plethora of sectors of the society. It is also concerned with the fact that as the country pushes for technological progress, advanced deep-fakes will be the fall out of such development and will be invading more sectors of the society as part of the technological proliferation process. India does not have a dedicated regulatory framework for addressing this challenge, and currently has to rely upon its existing legislation like the Information Technology Act, 2000 (ITA), aided by the general criminal provisions of the Indian Penal Code, 1860 (IPC). However, *vide* notification (S.O. 850(E)) by the Home ministry in the official gazette, July 1, 2024 marks as the effective date for

enforcement of the new panal code *viz*., the Bharatiya Nyaya Sanhita, 2023 (BNS)[50], replacing the IPC, about a sesquicentennial old law.[51]

Although the National Strategy and Responsible AI (NSAI) published by NITI Aayog in 2018 provided a basic framework for the responsible use of AI in India, it may not adequately cover concerns regarding deep-fakes and the subsequent requirement to safeguard personal privacy. A strategy formulated in 2019 to tackle the rapidly evolving and unpredictable digital landscape does not adequately address the issues posed by deep-fake technology in 2024.

Section 66E of the IT Act of 2000 covers deep-fake offences, such as unlawful recording, publication, or transfer of an individual's photographs through mass media platforms, which violates their privacy. This offence carries a potential three-year jail sentence or a fine of up to ₹2 lakh. Section 66D of the IT Act is also pertinent in this context. The aforementioned legislation enables the legal pursuit of individuals who employ communication devices or computer resources with the intent to engage in fraudulent activities or impersonate another individual. A conviction under this section can result in a three-year prison sentence and/or a ₹1 lakh fine. Deep-fakes call for stronger laws.[52] The Intermediary Rules of 2021, among several provisions, specifically deals with the issue of deep-fakes. Certain other relative legal provisions to safeguard against deep-fake can be traced under the Indian legal system as exhibited in the table below.

[50] Bharatiya Nyaya Sanhita, 2023, Law No. CG-DL-E-25122023-250883, Dec. 25, 2023, (India), https://www.mha.gov.in/sites/default/files/250883_english_01042024.pdf. (Last Accessed: 16 January, 2024).

[51] Bureau (2024). Three Criminal Laws To Be Effective From July 1. *The Hindu.* [online] 24 Feb. https://www.thehindu.com/news/national/three-newly-enacted-criminal-laws-to-come-into-effect-from-july-1/article67881602.ece. (Last Accessed: 25 February, 2024).

[52] Shankar, V. (2023). *Deepfakes Call For Stronger Laws*. [online] Business Line. https://www.thehindubusinessline.com/business-laws/deepfakes-call-for-stronger-laws/article67077019.ece. (Last Accessed: 19 January, 2024).

Table 3. India's Deep-fake Laws and Regulations

Sl. no.	Regulation	Year	Deep Fake Related Content
1.	The Information Technology Act	2000	Section 66 E – Punishment for violation of privacy; Section 66 D – punishment for cheating by personation by using computer resource; Section 67 – punishment for publishing or transmitting of material containing sexually explicit act, etc. in electronic form; Section 67 A- punishment for thr offence committed under Section 67. Section 67 B – punishment for publishing or transmitting of material depicting children sexually explicit act/pornography in electronic form.
2.	Indian Penal Code	1860	**Section 153** (s) and (b) (Inciting animosity among various groups based on religion, race, place of birth, place of residence, language, etc., and engaging in acts that undermine the preservation of harmony). **Sections 292** (sale etc. of obscene books, etc.) and **294** (obscene acts and songs) **Sections 465** (punishment for forgery) and **469** (forgery for purpose of harming reputation)[53] **Section 499** (Defamation) **Section 509** (An affront (insult), whether via words, gestures, or actions, that is

[53] PTI (2023). DCW Seeks Action Taken Report from Delhi Police On Rashmika Mandana Deep Fake Video. *The Hindu*. [online] 10 Nov. Available at: https://www.thehindu.com/news/cities/Delhi/dcw-seeks-action-taken-report-from-delhi-police-on-rashmika-mandana-deep-fake-video/article67521073.ece (Last Accessed 14 January, 2024)

			designed to demean the modesty of a woman.)
3.	Protection of Children from Sexual Offences Act, 2012 (POCSO)	2012	Sections 13,14 & 15 to protect the rights of women and children.
4.	Indecent Representation of Women (Prevention) Act	1986	Combats the inappropriate depiction of women.
5.	Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules	2021	Imposes obligations to expeditiously remove misinformation, failing which liability is imposed.
6.	Indian Copyright Act	1957	Section 51: "deemed infringement"
7.	Digital Personal Data Protection Act	2023	Data protection and privacy

Source – Compilation of Indian statutory provisions and Regulations

The existing regulatory framework pertaining to cyber offences facilitated by deep-fakes in India is insufficient in its ability to comprehensively and effectively tackle the issue.[54] The absence of explicit provisions poses challenges in effectively governing its utilization.[55] India possesses certain legal provisions that can potentially be utilized to tackle the issue of deep-fake technology. On the other hand, India is yet to witness a defining legal battle specifically centred on deep-fakes, but the potential for abuse is recognized by the legal community. However, it is imperative to develop more targeted legislation that specifically addresses the distinct issues presented by deep-fakes.[56] The Digital Personal Data Protection Act of 2023 (DPDPA) is a step in the right direction, as it is expected to address the technological ecosystem once implemented. At present, different

[54] *Supra Note 51.*
[55] Pandey, S., and Jadhav, G. (2023). *Emerging Technologies and Law: Legal Status of Tackling Crimes Relating to Deepfakes in India*. [online] SCC Blog. https://www.scconline.com/blog/post/2023/03/17/emerging-technologies-and-law-legal-status-of-tackling-crimes-relating-to-deepfakes-in-india/. (Last Accessed: 19 January, 2024).
[56] *Supra Note 51.*

provisions of the IT Act, 2000 and IT Rules 2021 can be deployed to commence legal proceedings against anyone involved in deep-fake cybercrimes within the jurisdiction of India.

Infact, the present plan of the Government is to notify specified amendments in the Intermediary Rules, 2021, aimed at defining deep-fakes, expanding the scope of "grievance" and notification to users about disallowed content,[57] in addition to the advisory issued to all intermediaries regarding AI generated prohibited content. [58]

VIII. Ethical Boundaries in Deep-fake Generation: Approaches and Technical Solutions

The prevailing techniques for detection primarily centre on uncovering the minute flaws introduced by algorithms, such as boundary artefacts like artificial corners of the lips, discrepancies in shading, or the presence of duplicated eyebrows. However, it's crucial to remember, though, that technology is always improving and therefore the vulnerabilities that exist today may potentially transform into advantages in the future. According to Professor Hany Farid, a renowned expert in the field of digital forensics and the creator of PhotoDNA, a cutting-edge technique designed to detect and prevent the dissemination of child pornography, opined that it is unlikely to achieve a level of forensic technology capable of definitively distinguishing between authentic and manipulated content for several decades.[59] To effectively deceive the system, one needs to incorporate techniques within deep-fake technology that exploit vulnerabilities in forensic systems.

Organizations that handle sensitive data or offer high-risk services are obligated to establish robust authentication systems in order to verify the authenticity of

---

[57] Agarwal, A. (2024). *Deepfake-Related Rules In 7-10 Days: Govt. To Tech Firms*. [online] https://www.hindustantimes.com/india-news/deepfakerelated-rules-in-7-10-days-govt-to-tech-firms-101704997856911.html (Last Accessed: 19 January, 2024).

[58] Press Information Bureau (Meity, Govt. of India) (2023). *Union Government Issues Advisory To Social Media Intermediaries To Identify Misinformation And Deepfakes*. [online] https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1975445#:~:text=The%20Minister%20added%20that%20Deepfakes (Last Accessed: 20 January, 2024).

[59] Farid, H. (2018). Reining in Online Abuses. *Technology & Innovation*, 19(3), pp.593–599. doi:https://doi.org/10.21300/19.3.2018.593. (Last Accessed: 22 January, 2024).

their transactions. Neglecting to adhere to this precautionary measure could potentially grant fraudsters entry and lead to unapproved financial activities. Therefore, it is imperative for organizations to develop more advanced detection techniques in order to enhance the overall welfare of all individuals. Additionally, policies should be de-signed to create an inclusive environment that supports the well-being of stakeholders, victims, and vulnerable populations, among others.

IX. Drawing the Line with Technology Support

Open-source software like DeepFace Lab and Faceswap allows anybody with enough time and cloud computing resources to deploy complex machine learning algorithms and graphics rendering approaches without programming skills. Because technology is advancing rapidly, experts worry that deep-fakes could soon become indistinguishable from true videos. Cryptographic methods have the capability to digitally sign or watermark not only textual content but also videos. The genuineness of a video can be readily ascertained if it has been digitally signed by a reputable entity.

China has implemented comprehensive regulations mandating that modified content must get the approval of the subject and be accompanied by digital signatures or watermarks. Additionally, deep-fake service providers are required to provide mechanisms for the purpose of countering false information. However, China encounters similar challenges that have impeded previous attempts to regulate deep-fakes. The individuals who exploit this technology to the greatest extent are often the most elusive, operating under anonymity, swiftly adapting, and disseminating their artificially generated content across unrestricted online channels. As anyone who has been the target of a deep-fake can attest, "no federal legislation presently criminalises deep-fake.[60] According to Brandie Nonnecke, a tech policy specialist and founding director of the CITRIS Policy Lab, the issue primarily rests with technology companies. Nonnecke suggests that these companies have the potential to establish self-governance mechanisms that would involve verifying an individual's consent for the utilisation of their facial features, name, and likeness. This proposal was

[60] D'Anastasio, C. and Alba, D., *Supra Note 34.*

shared with Bloomberg.[61] A victims' best hope for justice is for tech companies to "grow a conscience".[62] In other words, The optimal prospect for justice for victims lies in the development of ethical awareness within tech businesses.

X. Conclusion

Comparatively speaking, India requires prompt legislative determination and intervention to handle deep-fakes, even though such measures are also in its infancy in technologically advanced jurisdictions like the USA and EU. Given the complexity of India's socio-legal landscape, unregulated availability and use of such a technology can open a new portal of misery and trauma for a vast majority of the people in the country and should be treated with due concern by the policy makers in the country. However, after the recurring deep-fake incidents in 2023 beginning from the Rashmika Mandana deep-fake incident, India is acknowledging the risks posed are real and that symbolic treatment of the same will not suffice. As India has enacted the DPDP, a national law relating to protection of digital privacy which awaits enforcement, it is asserted that this is the right time and juncture to accord serious consideration to the risks posed by Deep-fakes on the societal and moral fabric of the nation and should be handled accordingly. According to Rajeev Chandrasekhar, the minister of information technology, the government is now drafting a regulatory framework with the intention of unveiling it in June or July 2024.[63] Like Europe, India should similarly commit funding and resources towards conducting research on modified or synthesized media in order to address and minimize the potential issues associated with it. India may consider adopting a strong law, relative to the AIA of Europe, to effectively address the issue of deep-fake in accordance

[61] Wu, T. (2023). *California Looks to Boost Deep-fake Protections Before Elections*. [online] https://news.bloomberglaw.com/artificial-intelligence/california-looks-to-boost-deep-fake-protections-before-elections. (Last Accessed: 25 January, 2024).

[62] Guglielmo, C. (2023). *AI and You: No Copyright for Nonhuman Creations, AI 'Boss From Hell,' Porn Deep-fakes*. [online] https://www.cnet.com/tech/computing/ai-and-you-no-copyright-for-nonhuman-creations-ai-boss-from-hell-porn-deep-fakes/ (Last Accessed: 25 January, 2024).

[63] Centre Working On Draft AI Regulation Framework: Three Things The Government Is Focussing On. (2024). *The Times of India*. [online] 21 Feb. https://timesofindia.indiatimes.com/gadgets-news/centre-working-on-draft-ai-regulation-framework-three-things-the-government-is-focussing-on/articleshow/107862302.cms (Last Accessed: 24 February, 2024).

with its current conditions, just as it did with the DPDPA influenced by GDPR. Nevertheless, a substantial amount of rigorous research is required to develop a legislation that is sufficiently successful in addressing the challenges posed by a vast population with a higher likelihood of deep-fake disruptions. It is also necessary to thoroughly explore complex topics such as AI that has been labelled. One of the ways this may be done is by utilization of a hybrid system comprising automated mechanisms and human moderators to detect and eliminate deceptive or inaccurate information. Furthermore, these platforms have the potential to engage in collaborative efforts with fact-checking organizations in order to authenticate the veracity of material disseminated on their platforms. Moreover, it is imperative to acknowledge the pivotal role that governments have in the regulation of disinformation dissemination.

Nevertheless, it is crucial to achieve a harmonious equilibrium between regulatory measures and the preservation of the fundamental right to freedom of expression. The complexity of effectively enforcing regulations on this type of content presents a significant challenge. There exists an urgent requirement for the development of a new ethical framework for the use of AI in the domain of political advertising and content dissemination on online platforms. Given the international nature of the matter under consideration, it is crucial that the suggested agenda receives backing from a thorough global consensus and subsequent execution. The extensive scope and rapidity of information propagation on social media provide challenges in identifying and remedying all occurrences of misinformation. Furthermore, the inherent subjectivity involved in the process of identifying the definition and boundaries of disinformation introduces an additional level of intricacy.