The Indian Journal for Research in Law and Management

Open Access Law Journal – Copyright © 2024

Editor-in-Chief – Dr. Muktai Deb Chavan; Publisher – Alden Vas; ISSN: 2583-9896

DEEP FAKE TECHNOLOGY: ANALYSIS OF LEGAL FRAMEWORK AND THE WAY FORWARD

~ Aldrin Kolakkal

INTRODUCTION

Deepfakes are a form of digital manipulation of audio or visual data. It is a form of cybercrime and often involves replicating an entities identity using various technological tools to disseminate false information. Creating deepfakes has never been more easier and prevalent as with the increase in digital penetration and easy access to Artificial Intelligence (AI), Photoshop, and Machine Learning (ML) software which are extensively employed to create convincing and authentic replica of videos and audio clips. Tweaking and manipulating existing images, videos and audio of individuals from social media and other platforms, cybercriminals produce content that is challenging to distinguish from reality in order to malign a person's character and spread false information.

Cyber criminals use facial mapping technologies to produce facial symmetry data set. They use Generative Adversarial Network (GAN technology) to swap the face of a person onto the face of another person. Apart from this, voice matching technology is used to accurately copy the user's voice. With all these layers of thoroughly edited changes; body movements, voice and facial expressions, it becomes really synchronous which makes it really hard to say if a digital media is fake or real. There are several reports which talk about the perils and dangers of deepfake technology in the modern era where technology has become really accessible and people often fall

prey to the ill ambitions and maligned objectives of cyber criminals. A report from the University College London (UCL) even went ahead to call deepfake technology one of the biggest threats of our time. The vastness, variety and veracity of crimes involving deepfakes make it really difficult to be regulated. But obviously this doesn't mean to call for an outright ban on the use of the technology as such would not be feasible; and the boons which the technology has to offer is something that can't be disregarded. In one such instance showcasing the marvel of this technology would be to talk about Malaria Must Die campaign where David Beckham (English Footballer) delivered an awareness program in 9 different languages. In many such arenas, deepfake technology is already being utilized , for e.g. Government schemes, interviews, and campaigns being launched in vernacular languages, etc. It becomes imperative for the legal framework to evolve to contain and regulate the ever growing technological advancements.

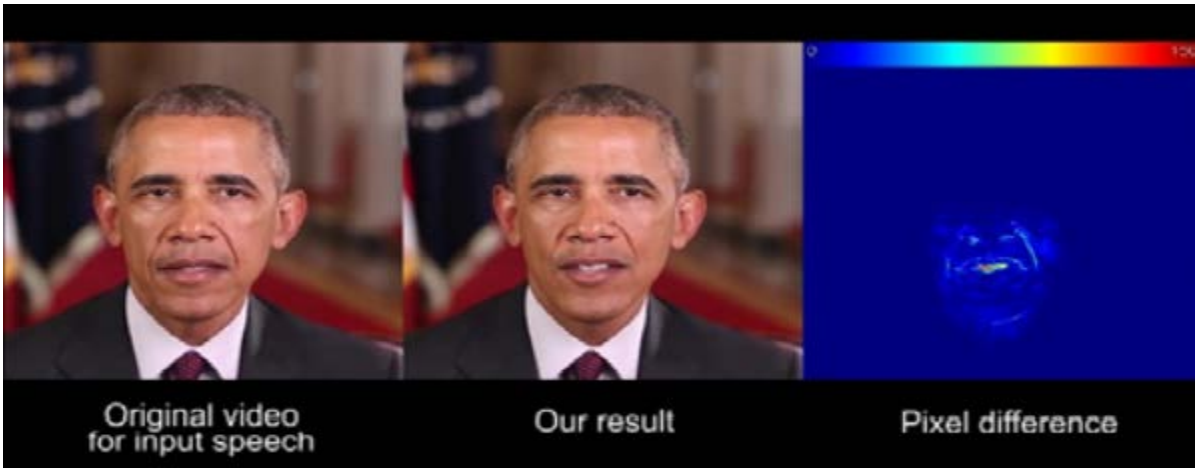
After thoroughly studying the Europol's Eurocrime Report several threats Deepfake Technology poses were understood and are listed as follows-

1. **Identity Theft**: This crime involves stealing someone's personal information, such as name, Social Security number, or credit card details, with the intent to misuse it for financial gain. Criminals might use this information to open new accounts, make unauthorized purchases, or even impersonate the victim.
2. **Cyberterrorism:** Acts like disrupting critical digital infrastructure, stealing sensitive data, or launching cyberattacks to cause widespread damage and threaten or inflict harm
3. **Cyberbullying:** Harassment on online platforms, social media, chat rooms, text messages, to spread rumors, defame victims, or cause emotional distress.
4. **Computer Intrusion:** Breaking into digital systems by unauthorized access and steal data, install malware, disrupt operations, or commit other malicious activities.
5. **Online Defamation:** Posting false or misleading information that damages someone's reputation and brings shame to a person in the eyes of the common man can be considered defamation and may have legal repercussions.
6. **Copyright Infringement:** With the rise of the internet, protecting intellectual property becomes more crucial. Copyright law grants creators exclusive rights to their original works, such as writing, music, or software. Distributing or modifying copyrighted material without permission from the copyright holder is a violation of this law.

CASES INVOLVING DEEPFAKE TECHNOLOGY-

1. During the ongoing Russia-Ukraine war, cybercriminals hacked a Ukrainian television channel and broadcasted a an altered video depicting the Ukranian President, Volodymyr Zelenskyy, surrendering. The fake video was created using deep fake hacking technology. After the invasion happened, another deepfake was released on Russian media, showing President Putin declaring full mobilization of Russian troops into the Ukraine offensive.

2. In 2018, filmmaker Jordan Peele and BuzzFeed CEO Jordan Peretti created a deepfake video to spread awareness about the prevalence of disinformation, specifically regarding the public's perception of political leaders. Peele and Peretti used free GAN tools with the help of editing experts to overlay Peele's voice and mouth over a pre-existing video of US President Barack Obama. In the video, Obama allegedly said, "We are entering an era in which our enemies can make it look like anyone is saying anything, at any point in time. Even if they would never say those things".



3. A few years ago, a deep fake video was reportedly created with Bharatiya Janata Party (BJP) leader Manoj Tiwari speaking in Haryanvi, Hindi and English. The video was

circulated via various WhatsApp groups ahead of the Legislative Assembly elections in Delhi in 2020.

4. Using new age technological digital tools and software it is possible to combine, or morph, or change the faces of the persons on important government documents like Identification Cards, driver's license, passport, etc. In the pictures attached below, the passport actually belongs to and the person(s) wanting to obtain a passport illegally. These techniques increase the chance that the photo in a forged document passes authorization, validity and identity checks including those using automated means (facial recognition systems).

The face in the middle of the image above is an example of a digitally manipulated facial image made using this 'morphing' method from the other two images. The images on the left and right are from The SiblingsDB, which contains different datasets depicting images of individuals related by sibling relationships.

WHY DEEPFAKE TECHNOLOGY POSE A THREAT TO LAW ENFORCEMENT-

With the growing evolution of Crime as a Service (CaaS) in parallel with such advancement in technologies is of increasing concern for law enforcement. CaaS criminals are using sophisticated skill and tools to break the law by selling access to the tools, technologies and knowledge to facilitate cyber and cyber-enabled crime. There has been a growing surge in CaaS much echoed during the COVID-19, resulting in the automation of crimes such as hacking and adversarial machine learning and Deep Fakes. Every time with the launch of a new technology many actors

have flagged the tendency of criminals to become early adopters of new technologies and bypass security mechanisms, laws and regulations for a longtime till the enforcement catches up with the technology. As a result, they are always one step ahead of law enforcement in their implementation, use and adaptation of these technologies.

Effect on law enforcement

Altered material on social media about events such as demonstrations may lead to police coming into action where it is not necessary, or in the wrong place. In police investigations, law enforcement may chase the wrong suspect of a crime when a deepfake version of the suspect fleeing a crime scene goes viral on social media, thereby giving the suspect the opportunity to get away. Using deepfakes, people could falsely portray police officers committing transgressions in order to discredit the police or even incite violence against officers.

Impact on the legal process

Rise of deepfakes would increase burden on the court to ascertain the authenticity and validity of evidence provided to the court. Usually audio-visual evidence whether the file is extracted from the phone of a suspect, downloaded from social media, or received from the CCTV system of a shop near the crime scene, the authenticity of the scene depicted is not usually questioned. With the rise of deepfakes, it will become increasingly important to scrutinize such content and verify if it is real or somehow artificially manipulated or generated. There would be a fundamental change to admissibility of such evidence.

HOW INDIAN LAWS ARE DEALING WITH CASES INVOLVING DEEP FAKE TECHNOLOGY

At present there aro no legislations which particularly deal with deepfakes in India. In most cases involving deepfakes, the available remedies at present which the victim can seek are under the Indian Penal Code 1860, Information and Technology Act 2000, Copyright Act 1957 and other remedies under tort law.

Deception and Identity Theft:

The Information Technology Act, 2000, serves as the cornerstone of India's cybercrime legislation. Section 66D specifically addresses the use of communication devices or computer resources with the intent to deceive. This offense carries a potential penalty of up to three years in prison and a substantial fine. Additionally, Sections 66 and 66-C9 of the same Act, along with Sections 420 and 468 of the Indian Penal Code, 1860, can be invoked depending on the nature of the deception and the associated harm.

Combating Misinformation and Cyber Terrorism:

Section 66-F12 of the Information Technology Act, 2000, and the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2022, aim to curb the dissemination of false information against the government and the incitement of hatred and disenchantment. These provisions play a crucial role in maintaining social harmony and protecting national security.

Safeguarding Sovereignty and Integrity:

For offenses that directly impact the sovereignty and integrity of India, maligning the stance of the government or threatening national security, the Indian Penal Code, 1860, provides potent legal instruments. Sections 121-A (waging war against the Government of India) and 124-A (Sedition Law) can be invoked in such cases, ensuring that the nation's security and communal harmony is not compromised.

Countering Hate Speech and Defamation:

Deepfakes, a potent tool for manipulation, can be utilized to spread hate speech and defamatory content, causing significant harm to individuals and society. To address this, Sections 153-A and 153-B (Speech affecting public tranquility) and Section 499 (defamation) of the Indian Penal Code, 1860, offer legal recourse.

Protecting Elections and Preventing Privacy Violations:

The Representation of the People Act, 1951, plays a vital role in ensuring the integrity of elections in India. Sections 123(3-A), 123, and 125, along with the Voluntary Code of Ethics for the General Election, 2019, provide a framework for tackling electoral malpractices and safeguarding the democratic process.

Furthermore, the Information Technology Act, 2000, through Sections 66-E, 67, 67-A, and 67-B, addresses the violation of privacy and the dissemination of obscene and sexually explicit content. Additionally, Sections 292 and 294 of the Indian Penal Code, 1860, and Sections 13, 14, and 15 of the Protection of Children from Sexual Offences Act, 2012 (POCSO), offer crucial protection for individuals, particularly women and children, in the digital space.

Intellectual Property and Data Protection:

The Copyright Act, 1957, safeguards intellectual property rights in the digital realm. Section 51 prohibits the unauthorized use of copyrighted material, ensuring that creators are duly recognized and compensated.

The upcoming Data Protection Bill, 2021, holds immense potential in further strengthening the legal framework for cybercrimes. Once enacted, this bill will provide comprehensive provisions to penalize the breach of personal and non-personal data, offering robust protection for individuals in the digital age.

MODERN DAY APPROACHES TO TACKLE DEEPFAKES-

There are many techniques, tools and technologies which are available to detect deepfakes. After studying the best practices of enforcement agencies and industry experts some approaches have been listed to make deepfakes easier to spot or to increase markers of authenticity.

Manual detection

While the sophistication of deepfakes is constantly evolving, trained human eyes can still identify inconsistencies in many cases. This manual detection process, however, is inherently labor-

intensive and can only be applied to a limited number of files at a time. Nonetheless, it remains a valuable tool, particularly in the early stages of deepfake analysis.

Several telltale signs can betray the artificial nature of a deepfake:

- **Facial inconsistencies:** Blurring around the edges of the face, unnatural eye movements (lack of blinking or inconsistent reflections), and discrepancies in hair, vein patterns, or scars can all point towards manipulation.
- **Background anomalies:** Inconsistent lighting, unrealistic shadows, or objects appearing or disappearing within the background can provide clues to the artificial nature of the content.

Automated detection

The ideal solution lies in the development of robust automated detection systems capable of scanning and analyzing digital content with high accuracy and efficiency. While achieving perfect accuracy may remain elusive, a reliable automated system would offer significant advantages over manual detection, particularly as the volume of deepfakes continues to grow.

Several organizations, including Facebook and security firm McAfee, have actively pursued the development of such software. These systems leverage advanced machine learning algorithms trained on vast datasets of real and deepfake content to identify subtle patterns and inconsistencies that may escape human scrutiny.

DETECTION TECHNOLOGIES THAT HAVE BEEN DEVELOPED IN RECENT YEARS-

Biological signals

One innovative approach focuses on analyzing subtle changes in skin color caused by blood flow variations within the face. These natural fluctuations are often difficult for deepfake models to accurately replicate, providing potential clues for detection. By analyzing video frames and employing advanced algorithms, researchers are exploring the feasibility of using this method to identify manipulated content.

Phoneme-Viseme Mismatches:

Another promising avenue involves analyzing the relationship between spoken sounds (phonemes) and the corresponding mouth movements (visemes). Deepfake models may struggle to perfectly synchronize these elements, leading to inconsistencies that can be exploited for detection. By comparing the extracted phonemes and visemes within a video, researchers are developing systems capable of identifying potential manipulations based on these discrepancies.

Facial Movement Analysis:

Facial movements and their correlation with head movements offer another avenue for deepfake detection. Each individual exhibits unique patterns in how their face moves in response to head gestures. By analyzing these correlations and comparing them to established baselines, researchers are developing systems that can distinguish between genuine and manipulated or impersonated content.

Recurrent Convolutional Models and Video Frame Analysis:

Videos are essentially composed of a sequence of individual images (frames). Deepfake detection techniques utilizing recurrent convolutional models (R-CNNs) analyze these frames for inconsistencies. By training these models on vast datasets of both real and deepfake content, researchers are enabling them to identify subtle discrepancies between frames that may be indicative of manipulation. This approach holds significant potential for automated detection, particularly as the volume of deepfake content continues to grow.

Expanding the Horizon:

The field of deepfake detection is constantly evolving, with researchers actively exploring additional avenues. These include:

- **Eye Gaze Analysis:** Examining eye movements and gaze patterns can reveal inconsistencies in deepfakes, as these elements can be challenging to accurately replicate.
- **Source GAN Detection:** Identifying the specific Generative Adversarial Network (GAN) model used to create a deepfake can provide valuable insights into its origin and potential manipulation techniques.

- **Blockchain Integration:** Utilizing blockchain technology to create tamper-proof records of content creation and provenance can offer an additional layer of verification and help combat deepfake manipulation.

PLATFORM RESPONSIBILITY IN TACKLING DEEP FAKE TECHNOLOGY-

Several major platforms have implemented policies aimed at tackling deepfakes. Meta (Facebook and Instagram) focuses on removing "edited media" where manipulation is not readily apparent and could mislead, particularly in videos. Similarly, TikTok bans "Digital Forgeries" that distort the truth of events. Reddit prohibits content that impersonates individuals or entities in a deceptive manner, explicitly including deepfakes intended to mislead or falsely attribute content. Youtube also has existing bans on manipulated media under its spam, deceptive practices, and scam policies.

However, these policies often rely on subjective assessments of "intent" to determine whether or not to remove a deepfake. This ambiguity presents challenges in enforcement, as defining and verifying intent can be highly subjective and context-dependent.

BEYOND CONTENT REMOVAL: MULTIFACETED APPROACH TO TACKLING DEEPFAKES

While content removal is a crucial step, online platforms can play a more proactive role in combating deepfakes. This includes:

- **User Education and Awareness:** Platforms can educate users on how to identify deepfakes, encouraging critical thinking and skepticism towards online content.

- **Transparency and Traceability:** Implementing mechanisms to track the origin and manipulation history of content can help identify the source of deepfakes and potentially hold creators accountable.
- **Collaboration with Law Enforcement:** Platforms should actively cooperate with law enforcement agencies to investigate and prosecute the malicious use of deepfakes.

The current patchwork of platform policies highlights the need for a more comprehensive and standardized approach to deepfake regulation. The proposed European Commission AI regulatory framework, while still under development, offers a promising step in this direction. This framework takes a risk-based approach, requiring deepfake creators to adhere to minimum standards, such as marking content as manipulated to ensure user awareness.

CONCLUSION

Tackling the challenge of deepfakes necessitates a collaborative effort from various stakeholders. Online platforms must go beyond content removal and actively promote responsible use of their technologies. Additionally, robust regulatory frameworks are essential to establish clear guidelines and hold creators accountable for malicious deepfakes. By working together, we can strive towards a safer digital landscape where users are empowered to critically evaluate online content and protected from the harms of deepfake manipulation.

REFERENCES-

1. FACING REALITY? LAW ENFORCEMENT AND THE CHALLENGE OF DEEPFAKES. (2022). In *Europol Innovation Lab*. Publications Office of the European Union. https://doi.org/10.2813/158794

2. Gamage, D., Sasahara, K., & Chen, J. (2021, January 1). *The Emergence of Deepfakes and its Societal Implications: A Systematic Review.* https://truthandtrustonline.com/wp-content/uploads/2021/10/TTO2021_paper_20.pdf

3. Wach, K., Duong, C. D., Ejdys, J., Kazlauskaitė, R., Korzyński, P., Mazurek, G., Paliszkiewicz, J., & Ziemba, E. (2023). The dark side of generative artificial intelligence:

A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, *11*(2), 7–30. https://doi.org/10.15678/eber.2023.110201

4. Jeong, D. (2020). Artificial Intelligence Security Threat, Crime, and Forensics: Taxonomy and open issues. *IEEE Access*, *8*, 184560–184574. https://doi.org/10.1109/access.2020.3029280

5. Ucl. (2022, May 6). *'Deepfakes' ranked as most serious AI crime threat*. UCL News. https://www.ucl.ac.uk/news/2020/aug/deepfakes-ranked-most-serious-ai-crime-threat

6. Parsons, J. (2022, March 4). Ukraine warns Russia may deploy deepfakes of Zelensky surrendering. *Metro*. https://metro.co.uk/2022/03/04/ukraine-warns-russia-may-deploy-deepfakes-of-zelensky-surrendering-16217350/

7. Yadlin-Segal, A., & Oppenheim, Y. L. (2020). Whose dystopia is it anyway? Deepfakes and social media regulation. Convergence, 27(1), 36–51. https://journals.sagepub.com/doi/10.1177/1354856520923963

8. Luciano Floridi. 2018. Artificial intelligence, deep-fakes and a future of ectypes. Philosophy & Tech-nology, 31(3):317–321

9. Michigan State University, MSU, 'Facebook develop research model to fight deepfakes', 2021,

   https://msutoday.msu.edu/news/2021/deepfake detection.

10. McAfee, 'The Deepfakes Lab: Detecting & Defending Against Deepfakes with Advanced AI',2020 https://www.mcafee.com/blogs/enterprise/securityoperations/the-deepfakes-lab-detecting-defending-against-deepfakes-with-advanced-ai.

11. Hongmei Chi, Udochi Maduakor, Richard Alo, andEleason Williams. 2020. Integrating deepfake de-tection into cybersecurity curriculum. In Proceed-ings of the Future Technologies Conference, pages588–598. Springer

12. U. A. Ciftci, I. Demir and L. Yin, "FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals," in IEEE Transactions on Pattern Analysis and Machine Intelligence, DOI: 10.1109/TPAMI.2020.3009287.

13. Agarwal, S. et al., 'Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches', 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020 https://www.ohadf.com/papers/AgarwalFaridFriedAgrawala_ CVPRW2020.pdf.

14. Agarwal, S. et al., 'Protecting world leaders against deep fakes', Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 38-45, 2019 http://www.hao-li.com/publications/papers/ cvpr2019workshopsPWLADF.pdf.

15. Cao, X. and Gong, N.Z., 'Understanding the Security of Deepfake Detection' ArXiv, 2021, accessed on 18 October 2022, https://arxiv.org/abs/2107.02045.

16. Meta, 'Manipulated media' https://transparency.fb.com/en-gb/policies/community-standards/manipulated-media/.

17. TikTok, 'Community Guidelines', https://newsroom.tiktok.com/en-us/combating-misinformation-and-election-interference-on-tiktok.

18. Becoming Human: Artificial Intelligence Magazine, ‘A Look at Deepfakes in 2022, https://becominghuman.ai/a-look-at-deepfakes-in-2020-13d3fe2b6ef7.

19. Reddit, ‘Updates to Our Policy Around Impersonation‘, 2020 https://www.reddit.com/r/redditsecurity/comments/emd7yx/updates_to_our_policy_around_impersonation.

20. Google Support, ‘Misinformation policies’, google.com/youtube/answer/10834785.

21. Arslan, F. (2023). Deepfake Technology: A Criminological Literature Review. *Sakarya Üniversitesi Hukuk Fakültesi Dergisi/Sakarya Hukuk Dergisi*, *11*(1), 701–720. https://dergipark.org.tr/en/pub/shd/issue/76836/1293642