IJRRAS

ISSN: 2349 - 4891

International
Journal of Recent Research and Applied Studies
(Multidisciplinary Open Access Refereed e-Journal)

A STUDY ON AI DETECTION IN DEEPFAKE - INDUCED FRAUD AND THE PROSPECTIVE EVOLUTION OF BHARATIYA NYAYA SANHITA,2023

Dr. SM. AZIZUNISAA BEGUM[1], Mr. SM. AYYUB[2], and Ms. K.KIRTHY[3]

[1] ASSISTANT PROFESSOR, SCHOOL OF LAW, VISTAS, CHENNAI
[2] RESEARCH SCHOLAR, NAGARJUNA UNIVERSITY, ANDHRA PRADESH
[3] LL.M, SCHOOL OF LAW, VISTAS, CHENNAI, kirthyg08@gmail.com

ABSTRACT

Deepfake technology leveraging advanced artificial intelligence (AI), has emerged as a significant threat, facilitating fraud and misinformation. With the increasing sophistication of Deepfakes, there is an urgent need to develop robust detection mechanisms and adapt legal frameworks to combat these threats effectively. This study aims to explore the effectiveness of AI detection methods in combating deepfake-induced fraud and examines the prospective evolution of the Bharatiya Nyaya Sanhita (BNS) to address these challenges. The objective is to provide a comprehensive approach to reducing deepfake-induced fraud and enhancing legal measures to combat such crimes effectively. The research employs an empirical method, analysing data from 205 samples. The study suggests several amendments to the BNS, including clear definitions of deepfake offences, criminalization of malicious deepfake activities, and stringent regulations for AI and digital platforms. Establishing specialised cybercrime units, protecting victims' rights, fostering international collaboration, and promoting technological innovation are identified as crucial steps. Educating the public on legal implications and incorporating digital literacy into educational curricula can further mitigate risks. Periodic review and updates of legal provisions will ensure the BNS remains adaptive to technological advancements. This study provides a comprehensive framework for addressing deepfake-induced fraud and suggests significant enhancements to the BNS to effectively combat such crimes.

[1] ASSISTANT PROFESSOR, SCHOOL OF LAW, VISTAS, CHHENNAI
[2] RESEARCH SCHOLAR, NAGARJUNA UNIVERSITY, ANDHRA PRADESH
[3] LL.M, SCHOOL OF LAW, VISTAS, CHENNAI, kirthyg08@gmail.com

KEYWORDS:

Deepfake Technology, AI Detection, Fraud Prevention, Bharatiya Nyaya Sanhita (BNS), Cybercrime, Digital Literacy

INTRODUCTION:

The advent of artificial intelligence (AI) has revolutionized numerous domains, one of the most significant being the detection and mitigation of deepfake-induced fraud. Deepfakes, synthetic media where a person in an existing image or video is replaced with someone else's likeness, pose a growing threat to cybersecurity, privacy, and trust. The proliferation of such technology has necessitated the development of sophisticated AI detection mechanisms to combat the misuse of deepfakes in fraudulent activities. The concept of deepfakes originated from advancements in artificial intelligence, specifically in the field of generative adversarial networks (GANs). GANs, introduced by Ian Goodfellow and his colleagues in 2014, consist of two neural networks—the generator and the discriminator—competing against each other to create increasingly realistic synthetic data. Initially, GANs were used for various benign applications, such as image synthesis, art generation, and data augmentation. The term "deepfake" itself emerged around 2017, when a Reddit user began posting videos that used GANs to swap faces in video clips, often involving celebrities. These early deepfake videos quickly gained notoriety for their realistic nature and potential for misuse. The technology rapidly evolved, with open-source tools like Fake App making it easier for individuals without deep technical expertise to create deepfake videos. The Ministry of Electronics and Information Technology (MeitY) has been at the forefront, launching programs to promote AI research and development. The establishment of the National Artificial Intelligence Portal and the drafting of the National Strategy on AI are significant steps towards creating a robust AI ecosystem in India. Additionally, the proposed Bharatiya Nyaya Sanhita, 2023, aims to address the legal implications of emerging technologies, including AI and deepfakes, ensuring that India's legal framework keeps pace with technological advancements. Several factors significantly influence the effectiveness of AI detection in combating deepfake-induced fraud. Technological advancements play a crucial role, as continuous improvements in AI algorithms and computing power enhance detection capabilities. Data availability is another critical factor; access to large datasets of both authentic and manipulated media is essential for training robust AI models. The regulatory framework also impacts the landscape, with legal measures and policies crucial in governing the use and misuse of deepfake technology. Public awareness is equally important; educating the public about the dangers of deepfakes and promoting digital literacy are essential steps in mitigating risks. Collectively, these factors shape the development and implementation of AI-based solutions to address deepfake fraud. The U.S. has seen significant investments in AI research from both the government and tech companies. Initiatives like the Deepfake Detection Challenge by Facebook and the Defense Advanced Research Projects Agency (DARPA) reflect a proactive stance in combating deepfakes. The EU emphasizes regulatory measures, with the General Data Protection Regulation (GDPR) addressing aspects of privacy and data protection. Additionally, the European Commission's efforts to develop ethical AI guidelines contribute to a comprehensive approach. Deepfake detection faces numerous technical challenges, such as the continuous improvement of deepfake generation techniques, which make synthetic media more realistic and harder to detect. In response, researchers are developing more sophisticated AI models that can analyze minute details in videos and images, such as subtle inconsistencies in lighting, reflections, and facial movements. Techniques like multi-modal analysis, which examines both visual and audio data, are becoming more prevalent. Moreover, the use of adversarial training, where detection models are trained alongside generation models, helps improve the robustness of detection algorithms. The deployment of AI detection systems raises ethical and privacy concerns. For instance, while detecting deepfakes is crucial for security and trust, it also involves the analysis of vast amounts of personal data, which could be misused or lead to privacy breaches. Ensuring that detection technologies are deployed responsibly, with respect for individuals' privacy rights, is essential. Additionally, transparency in AI processes and the establishment of ethical guidelines for the use of AI in media verification are important steps in addressing these concerns. Given the global nature of the deepfake threat, international collaboration is essential for developing effective detection solutions. Initiatives like the Global Partnership on AI (GPAI) and the Partnership on AI (PAI) facilitate collaboration between countries, promoting the sharing of knowledge, resources, and best practices. Developing international standards for AI ethics, data privacy, and security can help harmonize efforts across borders and enhance the overall effectiveness of deepfake detection technologies. The economic impact of deepfake technology extends beyond fraud, affecting various industries such as entertainment, media, and advertising. AI detection tools are increasingly being integrated into these industries to safeguard content authenticity and protect intellectual property. For instance, media companies use AI to verify the integrity of news footage, while advertising agencies employ detection technologies to ensure the authenticity of promotional content. The development of these tools also creates new economic opportunities in the fields of cybersecurity and digital forensics. To combat deepfake fraud effectively, it is essential to invest in educational and training programs for both professionals and the general public. Universities and research institutions are incorporating courses on AI ethics, cybersecurity, and digital media forensics into their curricula. Additionally, public awareness campaigns aim to educate people about the existence and risks of deepfakes, teaching them how to recognize manipulated media and encouraging critical thinking when consuming digital content.

OBJECTIVES:

- To analyse the effectiveness of AI-based deepfake detection techniques.
- To identify the challenges and development strategies for mitigating deepfake-induced fraud.
- To evaluate the adequacy of current legal frameworks related to Deepfakes detection.
- To suggest specific features or capabilities required for effective deepfake detection tools under the Bharatiya Nyaya Sanhita, 2023

LITERATURE REVIEW:

Diya Sarkar (2024) discussed 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.

Manupriya (2024) discussed companies and researchers working on tools to detect fake images and posts, for Deepfake has the potential of scamming people, creating disharmony, or triggering violence. While mechanisms for video, audio, and text alteration and editing have been around for a while, the widespread access to artificial intelligence (AI) tools has made it easier for anyone to alter or manipulate media.

Kaur (2024) explained the development of convincing fake content that threatens politics, security, and privacy. Training challenges include the need for many computational resources. It also addresses reliability challenges, including overconfidence in detection methods and emerging manipulation approaches. The research emphasises the dominance of deep learning-based methods in detecting deepfakes despite their computational efficiency and generalisation limitations.

Gayar (2024) explained the rise of deep fake technology has opened a new frontier in the digital world, enabling the creation of convincing synthetic video content. Deepfakes, artificial intelligence-based synthetic media where individuals in existing images or videos are replaced with someone else's likeness, have become increasingly prevalent. Deep fakes can now convincingly mimic facial expressions, lip movements, and even voices, making them virtually indistinguishable from real videos.

Pandey (2024) concluded Advancements in deepfake detection have proposed a combination of traditional techniques. With the evolution of technology and the low-cost barriers to entry, the advancement of deep fakes will progress with a rapid trajectory. As this evolves, future challenges in detection techniques will be required to adapt at the same level or greater to those technological advancements.

Borade & et.al (2024) explained that In recent years, there has been a pronounced escalation in cyber criminal activities, closely paralleling the rapid advancements in AI technology. The emergence of deepfake technology represents a formidable challenge to the maintenance of digital information integrity. To preclude the proliferation of erroneous information, bolster credibility, and uphold public trust, the implementation of detection systems is indispensable.

Imran & Tawde (2024) concluded that deep fake discovery is a quickly advancing field with noteworthy suggestions for society. Proceeded collaboration between analysts, policymakers, and the open is basic to create viable location methods, address lawful and moral concerns, and advance open mindfulness to moderate the potential hurts of deep fakes.

Patil & et.al (2024) analysed Deepfakes and Convolutional Neural Networks (CNN). Synthetic media generated by artificial intelligence (AI), pose a growing threat to the authenticity of visual content. Their ability to create realistic manipulations can lead to the spread of misinformation and compromise individual privacy. Define a function, test single image, to predict the class of a single image. Preprocess the image, make a prediction using the trained model, and display the result using Matplotlib.

Mubarak (2023) explains the broad implications of deepfakes in social, political, economic, and technological domains. State-of-the-art detection methodologies for all types of deepfake are critically reviewed, highlighting the need for unified, real-time, adaptable, and generalised solutions.

Naitali (2023) identified recent years have seen a substantial increase in interest in deep fakes, a fast-developing field at the nexus of artificial intelligence and multimedia.Deepfakes can be used for entertainment, education, and research; however, they pose a range of significant problems across various domains, such as misinformation, political manipulation, propaganda, reputational damage, and fraud. Therefore, deepfake detection techniques need to be constantly improved to catch up with the fast-paced evolution of generative artificial intelligence.

Piyush Jha (2023) analyses whether the Indian legal framework is adept at tackling the infringement of individual rights due to Deepfakes and argues that though the current laws impliedly prohibit the circulation of Deepfakes, there are a few challenges in the existing framework which can be addressed through suggested legislative amendments.

HYPOTHESIS:

ALTERNATE HYPOTHESIS:

There is a significant relationship between occupation and important steps to mitigate the risks of deep fakes.

NULL HYPOTHESIS:

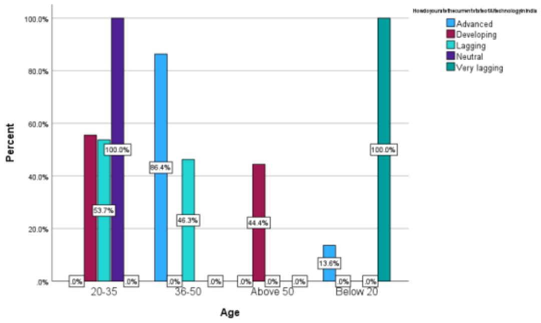
There is no significant relationship between occupation and important steps to mitigate the risks of deep fakes.

METHODOLOGY:

The study was based on an empirical method. The data was collected from the public by adopting the convenience sampling method and the sample size was 206. The responses were collected in Chennai. The independent variable in the analysis is age, gender, education, occupation and locality and the dependent variable is reliable on the statement. The tool used for the study was SPSS by using a structured questionnaire. Chi-square test is done to determine the hypothesis.

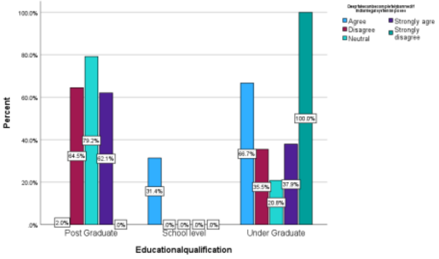
DATA ANALYSIS:

Figure 7



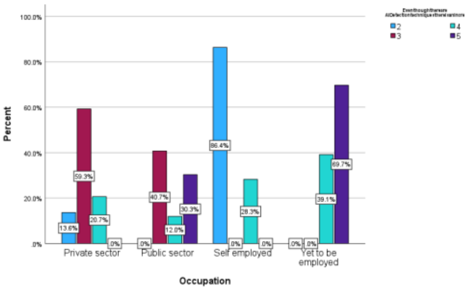
Legend: Figure 7 shows the age and current state of AI technology in India

Figure 8



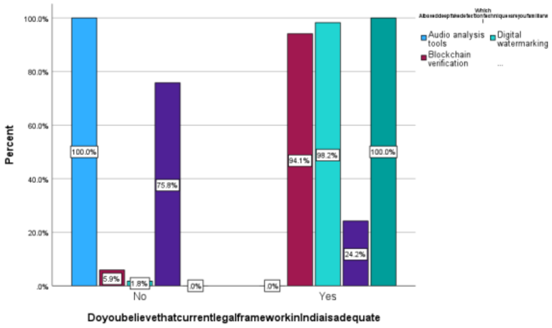
Legend: Figure 8 shows the education and deepfakes can be completely banned if Indian legal system implies proper penalties under BNS

Figure 9



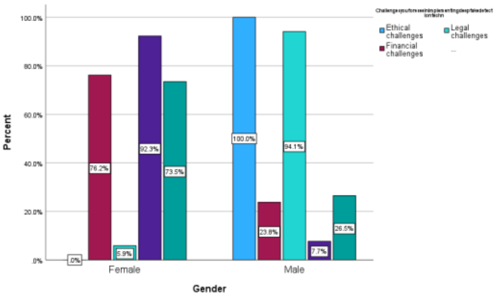
Legend: Figure 9 shows the occupation and rating on deep fakes in India.

Figure 10



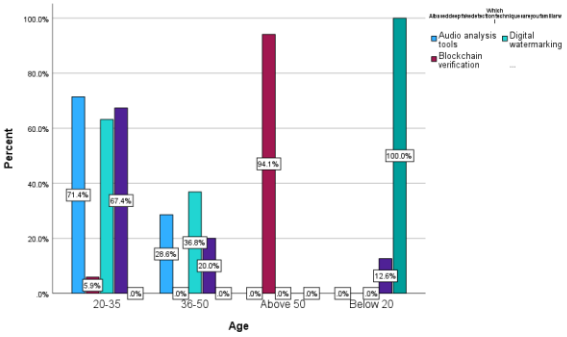
Legend: Figure 10 shows how many people believe the current legal and about techniques.

Figure 11



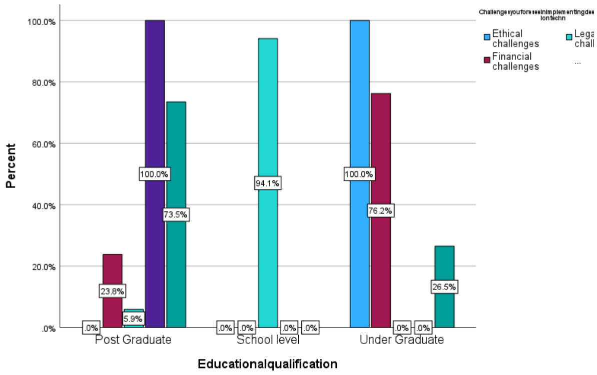
Legend: Figure 11 shows the gender and challenges.

Figure 12



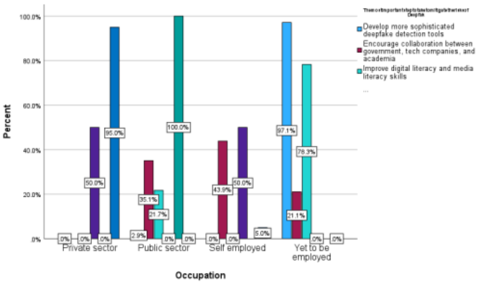
Legend: Figure 12 shows the age and about the deep fake techniques.

Figure 13



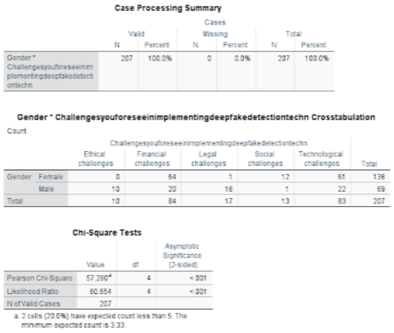
Legend: Figure 13 shows the education and challenges

Figure 14



Legend: Figure 14 shows the occupation and important steps to take to mitigate the risk of the deep fake.

Table 1



Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Challengesyouforeseeinimplementingdeepfakedetectiontechn	207	100.0%	0	0.0%	207	100.0%

Gender * Challengesyouforeseeinimplementingdeepfakedetectiontechn Crosstabulation

Count

		Challengesyouforeseeinimplementingdeepfakedetectiontechn					
		Ethical challenges	Financial challenges	Legal challenges	Social challenges	Technological challenges	Total
Gender	Female	0	64	1	12	61	138
	Male	10	20	16	1	22	69
Total		10	84	17	13	83	207

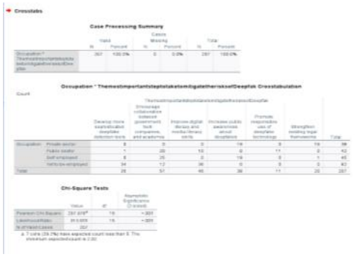
Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	57.280^a	4	<.001
Likelihood Ratio	60.654	4	<.001
N of Valid Cases	207		

a. 2 cells (20.0%) have expected count less than 5. The minimum expected count is 3.33.

Legend: Table one shows Chi square test between gender and challenges.

Table 2



Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Occupation * ThemostimportantstepstotaketomitigatetheriskofDeepfak	207	100.0%	0	0.0%	207	100.0%

Occupation * ThemostimportantstepstotaketomitigatetheriskofDeepfak Crosstabulation

Count

		ThemostimportantstepstotaketomitigatetheriskofDeepfak						
		Develop more sophisticated deepfake detection tools	Encourage collaboration between government, tech companies, and academia	Improve digital literacy and media literacy skills	Increase public awareness about deepfakes	Promote responsible use of deepfake technology	Strengthen existing legal frameworks	Total
Occupation	Private sector	0	0	0	19	0	19	38
	Public sector	1	20	10	0	11	0	42
	Self employed	0	25	0	19	0	1	45
	Yet to be employed	34	12	36	0	0	0	82
Total		35	57	46	38	11	20	207

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	[illegible]^a	15	<.001
Likelihood Ratio	[illegible]	15	<.001
N of Valid Cases	207		

a. 7 cells (29.2%) have expected count less than 5. The minimum expected count is 2.02.

Legend: Table 2 shows the Occupation and the most important steps risk of Deep fakes.

RESULTS:

Figure 7 shows 99% of respondents opted for neutral opinion on the current state of AI technology in India. **Figure 8** shows that 66.7% of respondents agree that if Indian legal system imposes stricter penalties, then the deep fix can be reduced. **Figure 9** shows that 86.4% of respondents opted at the rate of to that, even though there are AI detection techniques Deep fakes are increasing. **Figure 10** shows that 98.2% of respondents opted for facial recognition as a well-known dipstick. **Figure 11** shows that 92.3% of respondents opted for technical challenges as a challenge, this force in on the deep fake technology. **Figure 12** shows that at 94.1% of respondents opted for block chain verification as a familiar deep fake. **Figure 13** shows that 99% of post graduates opted for technical challenge is a major thing to implement detection. **Figure 14** shows that 97.1% of respondents opted for develop more sophisticated defect detection tool, Aisa most important steps to mitigate the risk of defect. **Table 1** shows that there is a significant relationship between gender and challenges. **Table 2** shows that there is a significant relationship between occupation and important steps, to mitigate the risk of deepfakes.

DISCUSSION:

In **Figure 7**, most of the respondents opted for a neutral opinion because there is a beach of definitive stance on the current state of AI technology in India.In **Figure 9,** most of the respondents indicated that deepfakes are increasing despite AI detection techniques because there is a beach between the effectiveness of current detection methods and the evolving sophistication of deepfake technology. most of the respondents agree that imposing stricter penalties could reduce deepfakes because they believe there is a beach in the legal system's severity.In **Figure 10**, most of the respondents opted for facial recognition as a well-known dipstick because it addresses the beach in commonly understood and trusted methods for identifying deepfakes. In **Figure 11**, most of the respondents identified technical challenges as a significant issue because there is a beach in overcoming technological hurdles in deepfake detection.In **Figure 12,** most of the respondents opted for blockchain verification as a familiar method because it addresses the beach in trust and verification processes.In **Figure 13**, most postgraduates opted for technical challenges as a major issue because there is a beach in the implementation of effective detection methods. In **Figure 14,** most of the respondents believe developing more sophisticated detection tools is crucial because there is a beach in the capabilities of current tools. **Table 1** shows that Null hypothesis is rejected and alternate hypothesis is accepted. **Table 2** shows that Null hypothesis is rejected and alternate hypothesis is accepted.

LIMITATIONS:

One of the major limitations of the study in the sample frame. There is a major constraint in the sample frame as it is limited to a small area. Thus, it proves to be difficult to extrapolate it to a larger population. Another limitation is the sample size of 206 which cannot be used to assume the thinking of the entire population in a particular country, state, or city.

SUGGESTIONS:

Public awareness campaigns, digital literacy programs, and the use of AI-powered detection tools can empower individuals to identify and report deepfakes. Additionally, amendments to the BNS should include clear definitions and classification of deepfake offenses, criminalization of malicious deepfake creation and distribution, and implementation of regulations for AI and digital platforms. Establishing specialized cybercrime units, protecting victim rights, promoting international collaboration, and fostering technological innovation are crucial. Educating the public about legal implications and including digital literacy in curricula can further mitigate risks. Periodic review and updates of legal provisions will ensure the framework keeps pace with technological advancements. These combined efforts can significantly reduce the incidence of deepfake-induced fraud and enhance legal measures to combat such crimes effectively.

CONCLUSION:

The study of AI detection in deepfake-induced fraud underscores the dynamic and rapidly evolving nature of both the threats posed by deepfakes and the technologies developed to counteract them. From their origins in generative adversarial networks to the sophisticated, hyper-realistic media manipulations of today, deepfakes have continually pushed the boundaries of what is possible with synthetic media. This evolution has prompted significant advancements in detection methodologies, leveraging deep learning, multi-modal analysis, and adversarial training to keep pace with the growing sophistication of deepfakes. Government initiatives, such as India's proposed Bharatiya Nyaya Sanhita, 2023, highlight the crucial role of legislative frameworks in addressing the legal and ethical challenges associated with deepfake technology. Public awareness and education, coupled with robust regulatory measures, are essential components in mitigating the risks of deepfake-induced fraud. International collaboration and the development of global standards will be pivotal in ensuring a coordinated and effective response to the deepfake phenomenon. As AI technology continues to advance, it is imperative that detection capabilities and legal frameworks evolve in tandem, fostering a secure digital environment where the integrity of information and personal privacy are protected. Ultimately, the ongoing battle between deepfake creators and detectors exemplifies the broader challenges and opportunities of the digital age, necessitating continuous innovation, vigilance, and cooperation across all sectors of society.

REFERENCES:

1.Diya Sarkar, July 2024, Combating Deep Fakes in India- An Analysis of the Evolving Legal Pardigm and its challenges, 15(1).

2. Manupriya, March 2024, Tech tools to fight Deepfakes, Vol 3, Iss 2, Deep Fake artificial intelligence Generative AI Technology.

3. Kaur, A., Noori Hoshyar, A., Saikrishna, V. *et al.* Deepfake video detection: challenges and opportunities. *Artif Intell Rev* **57**, 159 (2024).

4. Gayar & Sweidan, 2024, A novel approach for detecting deep fake videos using graph neutral network, Journal of Big data.

5. Pandey, International Journal of Research Publication and Reviews, Journal homepage: www.ijrpr.com ISSN 2582-7421, Deepfakes Detection and Prevention.

6. Borade & et.la, Deepfake Technology and the Detection Methods, Journal of Emerging Technologies and Innovative Research, Volume 11, Issue 4.

7. Imran & Tawde, Deepfake Detection: A Literature Review, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 11 Issue: 03 | Mar 2024

8. Patil, Vanmali, Raut, Kazi, Deep Fake Image Detection using Artificial

Intelligence2024 IJCRT | Volume 12, Issue 4 April 2024 | ISSN: 2320-2882.

9. Mubarak, Tariq, Omar Alshaikh, Isa Inuwa- Dute, January 2023, A survey on the detection and impacts of Deepfakes in Visual, Audio and Textual formats, IEEE Access, PP(99): 1-1.

10. Naitali & et.al, Deepfake attacks: Generation, Detection, Datasets, Challenges and Research Directions, 2023, 12(10)

11. Piyush Jha, 19 April 2023, Detecting and Regulating Deepfakes in India: A Legal and Technological Conundrum

12. Arash Heidari, Nima Jaaferi Navimipoud, 20 November 2023, Deepfakes detection using deep learning methods: A systematic and comprehensive review.

13. Sudeep Tanwar & et.al, December 2023, Deepfakes detection and Detection Case study and Challenges, IEEE Access 11(2023): 14296-143323.

14. Gourav Gupt, Kiran Raj, Manish Gupta, Tony Jan, A comprehensive review of DeepFake Detection using Advanced Machine Learning and Fusion Methods, *Electronics* **2024**, *13*(1), 95.

15. Rimsha, 8 May 2023, Deep fake detection and classification using error level analysis and deep learning, Article number 7422 (2023).

16. Harini P & et.al, Deep Fake Detection using Deep Learning, International Journal of Science, Engineering and Technology, 2023, 11:5.

17. Malik & et.al, Deep Fake Detection for Human Face Images and Videos -A survey, February 2022.

18. Guhagarkar, Sanjana Desai, Swanand vaishyampayan, Ashwini Save, Deep Fake Detection Technique- A Review,International Journal for Research and Innovation, Volume 1, Issue 4 (2021).

19. Karthik, Review of Deepfake Detection Techniques, International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181 Vol. 10 Issue 05, May-2021

20. Aarti Karandikar & et.al, International Journal of Advanced Trends in Computer Science and Engineering, 9(2), March - April 2020, 1311 – 1315