DEEPFAKES IN THE AGE OF TECHNOLOGY: ISSUES, ABUSES, AND CYBER LAWS

Name of Author: Jashodha Barai / Jyoti Yadav
MSc Information and Cyber Security Student
Department of Information Technology
Guru Nanak Khalsa College, Mumbai, India

Abstract: AI has brought about significant developments and advancements in the field of digital communications and media. One of those innovations in the AI field is deepfake technology that creates highly realistic and, at the same time, fabricated audio visual material using deep learning techniques. There are plenty of beneficial applications of deepfakes including entertainment, accessibility, education, and much more, but the malicious use of this technology has created numerous problems for the fields of cybersecurity, privacy, law, society, ethics, and others. Some popular use cases include misinformation, identity fraud, damage to someone's reputation, politics influence, scam, pornography, and more. Therefore, deepfake detection has been one of the crucial aspects of research and development in the fields of digital forensics and trustworthy AI. This paper aims to provide a broad overview of deepfakes, deepfake detection technologies, and the legal implications. The paper reviews different traditional machine learning approaches, deep learning algorithms, multimodal approaches, and XAI approaches, among other things, mentioning limitations in each case. Apart from that, the legal perspective on deepfakes in this paper will be discussed with regards to the Indian Legal Framework like the IT Act, 2000, IT Rules, 2021, DPDPA, 2023, and Bharatiya Nyaya Sanhita, 2023.

Keywords : *Detection of deepfake technology, IT Act 2000, cyber law, artificial intelligence, Bharatiya Nyaya Sanhita 2023, India*

I. INTRODUCTION

The use of artificial intelligence technology has significantly improved the speed, cost-efficiency, and realism of producing digital content. Among various types of fake content, one of the most impactful ones is deepfake media, which refers to synthetic or altered videos created via deep learning techniques like autoencoders, generative adversarial networks (GANs), diffusion models, and neural voice generators [1], [2]. The problem with deepfakes is not only their high degree of realism but the ability of these videos to mimic the appearance, voice, actions, and facial expressions of a particular person.

At first, deepfake videos were seen mostly as an experiment or entertaining novelty for users. However, the technology rapidly gained traction due to its potential to cause harm to people. Currently, deepfakes are widely employed for impersonating someone, conducting scams, causing financial losses, defamation, political misinformation, creating pornographic content,

identity theft, and spreading disinformation [3], [4]. In many situations, the effect of a deepfake video is instantaneous and viral, making the removal of misleading information impossible once it is distributed across different social media platforms.Technically speaking, the detection of deepfakes has emerged as one of the most pressing issues in digital forensics. Researchers have devised a variety of approaches to detecting deepfakes using visual artifacts, inconsistencies in facial features, frequency domain properties, biological information, temporal dynamics, discrepancies between audio and video streams, and multimodal reasoning [1], [5], [6]. However, modern literature highlights the fact that while detection systems may perform well on benchmark data sets, their performance tends to falter during out-of-domain testing, especially for novel generation techniques and compressed social media posts [2], [7], [8].

Legally speaking, deepfakes defy existing laws since they encompass various crimes simultaneously such as privacy violation, defamation, obscenity, impersonation, fraud, misinformation, intellectual property violations, and cyber harassment. There is currently no standalone 'deepfake law' in India. Rather, the enforcement process depends upon several different pieces of legislation, such as the Information Technology Act of 2000, intermediary guidelines for due diligence, data protection laws, and general criminal laws [9]–[12]. Moreover, advisories issued by the government in 2023 acknowledged the rising threat of misinformation generated through artificial intelligence and instructed intermediaries to take proactive steps against malicious deepfakes [13], [14].

This paper adopts an inter-disciplinary and holistic approach to the subject matter. It does not restrict itself either to the technological issues involved or to purely legal questions. Instead, it attempts to explore both aspects of the topic – the technical one, which helps in detecting deepfakes, and the legal one, which is applicable even when detection is difficult or questionable.

The key contributions of this paper include the following:

1.To outline the development and functioning of deepfake technology.

2.To discuss different techniques of deepfake detection – traditional, deep learning-based, multimodal, and explainable.

3.To analyze the legal consequences of deepfake technology with regards to cyber laws.

4.To highlight the shortcomings of India’s existing legal system and regulations in this regard.

5.To provide a way forward for India.

II. RELATED WORK

Detection techniques for deepfakes have progressed significantly over the last few years, whereby there has been a significant evolution in the type of literature available in terms of deepfake detection, from simple detection based on artifacts to more advanced systems like multimodal and explainable deepfake detectors. Previous studies were mainly concerned with issues related to inconsistencies, unusual blinking rate, facial warpings, or blending artifacts. However, as the generation of deepfakes evolved with improvements in the underlying generative models, the performance of these approaches reduced significantly [15], [16]. The current surveys have revealed that contemporary detection techniques make heavy use of deep learning networks, transformers, frequency domain-based analysis, and multimodal fusion [1], [2], [5], [7]. For example, according to Soundarya and Gururaj's 2026 review paper, the recent trend in deepfake detection is towards robustness and interpretability as opposed to being

solely accuracy-driven [1].The research focus has not been limited to detecting either the image or the video but has evolved into audio-visual or multimodal detection since many of the modern day deepfakes involve manipulation of both types of media together [6], [7], [8]. The reason behind this approach is that a good deepfake involves synchronized facial movements, the generation of voice, and background setting, hence requiring the use of multimodal detection techniques which can recognize differences in synchronization of speech with lip movement and narration with visuals.Another prominent development relates to explainable detection. In legal contexts and forensics, simple outputs like "fake" and "real" are insufficient. Experts require explanations that are clear and defensible. Modern solutions include mechanisms for saliency maps, region-wise explanations, and human-readable rationales [8]. Explainability is crucial in cases where digital evidence can be questioned in courts.

Legally speaking, researchers have begun to emphasize that deepfakes are not only an issue of technology but also one of privacy, consent, non-consensual sexualized images, fraud, election security, and platform governance [4], [17], [18]. In India, scholars concur that despite existing legislation being able to tackle some aspects of deepfakes indirectly, there is no specialized legislative framework yet [9] – [14].

Hence, the current literature provides two critical insights into the problem: deepfake detection techniques are getting better, yet they are far from being perfect. Meanwhile, the response of legal systems to deepfakes is reactive, not anticipatory. Both of these insights provide the foundation for the following research paper.

III. UNDERSTANDING DEEPFAKE TECHNOLOGY

A. Meaning and Concept of Deepfakes

Deepfake is a portmanteau of two words, "deep learning" and "fake," referring to artificial intelligence-generated or manipulated media content that looks like a replica of an actual human, object, event, or setting to a human viewer [1], [2]. Deepfakes may manifest themselves in various forms such as face-swap manipulation, where one person's face is swapped with someone else; facial re-enactment, changing facial expressions or lip movements; voice-cloning, duplicating one's speaking style and tone; text-to-image/text-to-video creation; generating synthetic avatars/talking heads; and also multi-modal manipulation involving videos, audios, and semantic aspects of them altogether. The main threat posed by deepfakes arises from their potential to create extremely realistic fake evidence. Considering the fact that people face difficulties in identifying fake elements in the digital age anyway, deepfakes can capitalize on human emotions, trust, and virality of digital platforms for the spread of misinformation and misuse purposes.

B. Core Technologies Behind Deepfakes

Initial deep fake techniques employed autoencoders that learned latent representations of faces and generated synthetic images based on such representations [15]. Further development resulted in GANs since they could enable competition between generator and discriminator networks, resulting in realistic content generation [16]. Modern synthetic media is created using the techniques like diffusion models, transformers, and large multimodal generative models [1], [3].

Voice cloning requires the use of embeddings, text-to-speech synthesis pipeline, and vocoder in order to replicate tone, pitch, rhythm, and style of speech [6]. Combining it with lip-sync and facial animation enables creating highly convincing videos.

C. Legitimate and Illegitimate Uses

The deepfakes themselves are not illegal, as they also have multiple valid uses in various industries. For instance, deepfake technologies may be used positively in the film industry, voice dubbing, educational purposes, reconstruction of history, accessibility, and creation of avatars. Therefore, in the light of these examples, one can say that deepfakes are not inherently bad technologies and may even be useful.Nonetheless, the most significant ethical and legal problem lies in their abuse. The increasing use of deepfakes for criminal and malicious purposes is seen in the creation of non-consensual pornography, dissemination of propaganda and misinformation, fraudulent activities, such as CEO and executive impersonation schemes, identity theft, defamation, and even falsification of documents for legal disputes and litigation cases. As the generative AI technologies become less costly and user-friendly, the creation of malicious deepfakes is likely to become increasingly prevalent, which poses serious problems regarding cybersecurity, privacy, and enforcement of cyber laws [3], [4].

IV. DEEPFAKE DETECTION METHODS

A. Traditional Forensic and Machine Learning Approaches

The earliest approaches for detecting deepfakes were mainly aimed at spotting visual artifacts associated with the use of machine learning-based image synthesis models. For instance, they may involve unusual eye blinking behavior, unnatural-looking skin texture, abnormal edge blending, color discrepancies across the face boundaries, inconsistency in lighting or shadow effects, and head orientation anomalies. Such manually crafted forensic clues were then evaluated using conventional machine learning algorithms like SVMs, random forests, and logistic regression to classify the data as either real or fake [15]. This approach was quite successful at the initial stage of deepfake development since the generated videos had visual flaws that could easily be detected using these algorithms. However, the accuracy of such conventional detectors was greatly compromised as deepfakes became increasingly realistic.

B. Deep Learning-Based Detection

The detection of deepfakes in modern times is dependent on state-of-the-art deep learning models like CNNs, RNNs, 3D CNNs, transformers, and combinations of different architectures [1], [5]. Unlike previous approaches to deepfake detection, where the algorithms were trained to detect manipulation based on artifacts seen by the naked eye, deep learning models detect more complicated features and representations learned straight from large training data sets, allowing them to discover signs of manipulation that are less visible. The signals used by deep learning for detecting manipulation include spatial texture anomalies, inconsistencies in the frequency domain, temporal instability between video frames, physiological anomalies such as pulse, eye movement, or microexpressions, and even compression artifacts which can be indications of deepfake creation. Review articles have shown that while deep learning models perform well in benchmark datasets, they fail in real-world scenarios where they face new types of deepfakes [1], [2], [7].

C. Frequency-Domain and Signal-Based Detection

A number of studies have found that traces left behind in frequency space by synthetic media are not clear in pixel space [5].

This is particularly relevant considering that even visually realistic fakes could possibly have traces of non-humanly created regularities in their textures and distributions. Nonetheless, frequency-domain approaches could prove to be vulnerable when new generations of generators become adept at obscuring any frequency-space traces.

D. Temporal and Biological Signal Analysis

Video deepfakes tend to have trouble precisely replicating temporality between consecutive frames. This leads to the possibility for the development of more sophisticated methods of detection. Consequently, many of the currently available deepfake detectors for videos take into account the coherence of frames, head motion dynamics, eye closure and blinking rates, consistency of facial musculature, and rPPG data, which are utilized to estimate physiological processes, such as pulse rate, using facial videos. Among all the methods, biological pattern-based detection can be viewed as especially promising since natural physiological patterns have a lot of complexity, making their artificial reproduction rather complicated. Nevertheless, due to various limitations, some of these methods cannot always guarantee precise results, failing when faced with low-quality, short, or highly-compressed videos [2].

E. Audio Deepfake Detection

However, deepfake technology is also actively used in fraud and scams for social engineering as well as impersonation via voice synthesis. In order to discover fakes, scientists concentrate on various acoustic and linguistic features, which may be used for discovering any manipulation. Those include spectrogram distortions, prosodic discrepancies, unusual phoneme transition, breath and pause patterns, and inconsistent speaker embeddings that could indicate a mismatch with the characteristics of a natural speaker's speech. Such a strategy is based on discovering those traces, which speech synthesis software cannot reproduce. The latest research from NIST also emphasizes the critical need for evaluating evidence generated using artificial intelligence techniques. Moreover, the difficulties with development of trustworthy forensic systems are also indicated [19].

F. Multimodal Deepfake Detection

One of the most exciting directions in the latest research in this area is the multimodal approach to the detection of deepfakes, which is based on several sources of evidence rather than on one. The idea is to integrate a variety of complementary sources of evidence, such as facial visual cues, lip-syncing features, the speech-mouth relationship, semantic consistency of speech and lip movements, and cross-modality fusion of visual-auditory signals. Multimodal approaches appear to be highly effective because of the use of complementary information, which allows detecting certain inconsistencies between the various features and signals. It was demonstrated that the application of multimodal detection models results in superior outcomes compared to those obtained from monomodal models because they utilize more information for the detection process [6], [8], [20]. Nevertheless, the models under discussion face some problems. One of them is cross-dataset fragility, which leads to significant drops in the accuracy of predictions if the detector is applied to the dataset it was not trained on [7], [21].

H. Key Technical Limitations

Although there have been advancements in the development of deepfake detection techniques, several limitations exist which hinder their applicability in the real world scenario. For instance, current detection models exhibit poor generalization on data that they have never seen before. They may fail because of issues such as compression and scaling of videos. The process of generating deepfakes continues to evolve and become harder to detect by conventional approaches. Another major limitation of deepfake detection methods involves their susceptibility to adversarial attacks designed to bypass their functionality. Many machine learning based approaches lack forensic interpretability making it difficult to explain results generated by detection models. Lack of evidence for real world cases remains an issue, and most researchers rely on bench mark datasets. Bias towards benchmarking is another problem since deepfake detection models may only perform well within artificial environments. Therefore, despite obtaining high detection accuracy, one cannot treat deepfake detection systems as perfect or flawless. Deepfake detection can be treated as forensic probability evidence.

V. CYBER LAW IMPLICATIONS OF DEEPFAKES

Legal issues arise from the ability of deepfakes to concurrently infringe several rights and duties. Unlike typical edited content, deepfakes may create a fabrication regarding the identity, words, deeds, or approval of a particular person. These challenges include matters of attribution, liability, evidence, and accountability for platforms.

A. Breach of Privacy and Consent

Face, voice, and appearance of any person are significant components of their dignity and identity. If recreated without permission, especially through deception or sexually explicit content, the person may suffer psychological harm, defamation, and enduring online presence. Pornographic deepfakes constitute one of the most injurious instances of deepfake misuse since it is employed as a tool against the subject's autonomy [4], [17].

B. Slander and Reputation Damage

In a deepfake video, the subject may be presented doing or saying something defamatory. Even after being proved false, the reputation harm can remain. Legal avenues for defamation can be employed, although the enforcement process becomes challenging due to rapid dissemination via various platforms, anonymized accounts, encryption services, or foreign territories.

C. Frauds and Impersonations

Voice-cloning and video-cloning deepfakes are commonly used for phishing, business frauds, and identity theft. Individuals can clone executives, relatives, and government officials in order to prompt swift actions or money transfers [3], [4]. Therefore, besides being a form of misinformation, deepfakes can also be referred to as cybercrime.

D. Threats to Democracy

Deepfakes may affect democracy by influencing the attitude of the people using misleading videos and speeches that support or undermine particular individuals. When elections come,

even small amounts of misleading information may affect the voting behavior of people. The effects will certainly be amplified by the presence of algorithms in social media platforms.

E. Evidence and Forensic Issues

The most troubling legal issue posed by deepfakes, perhaps, is their threat to evidence authenticity. As long as digital materials like audio and video can be cloned, there might be doubts about their reliability on the part of the justice system, investigators, journalists, and ordinary people. This is referred to as "liar's dividend," which suggests that any evidence may be dismissed merely because deepfakes are present [18].

VI. Legal System In India

At this point, there are no particular laws that have been put in place in India to counter deepfakes. The strategy adopted is through the use of various laws.

A. Information Technology Act, 2000

Information Technology Act, 2000 has remained the cornerstone cyber law for tackling all sorts of cybercrimes and violations in India [9]. Though the Act does not provide any specific section for deepfakes, some sections of the Act may prove useful based on the type and nature of misusing deepfakes. For instance, Section 66C would apply where there is identity theft while Section 66D would apply in case of cheating by personation through computer resources, which holds special relevance in scams associated with impersonation via audio and video editing. Also, Section 67 deals with the publication or transmission of obscene material while Section 67A relates to sexual offenses. As such, Sections 67 and 67A would be highly relevant when deepfake is used in an abusive manner. Moreover, Section 69A allows blocking the information made available on computers while Section 79 grants immunity from liability to intermediaries provided they have exercised due diligence.

B. Information Technology Rules, 2021

According to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, intermediaries owe their due diligence towards preventing unlawful content [10]. Rule 3(1)(b), which is referenced frequently in government advisories, bans hosting and sharing of defamatory, obscene, privacy-invasive, impersonating or any other unlawful content [13], [14].

This is important because most of the dangerous deepfake content falls under these prohibited categories, even if the term "deepfake" itself does not appear in the statute.

C. 2023 Government Advisories on Deepfakes

In November and December 2023, the Government of India released public advisories directed to intermediaries about the need to take proactive measures in dealing with misleading and deepfake content [13], [14]. The 2023 November advisory mentioned the importance of taking down the content in accordance with the time limits set in the IT Rules, including removal in 36 hours where applicable [14]. In the 2023 December press information bulletin, there was a mention of AI-driven misrepresentations and deepfakes [13].

The significance of these advisories is the fact that the government recognizes deepfake as an important issue despite lacking a specific legislation for deepfake content.

D. Digital Personal Data Protection Act, 2023

The Digital Personal Data Protection Act, 2023 (DPDP Act) also offers another privacy angle [11]. As deepfakes are usually generated using one's personal image or voice or biometrics-

like identifier, the unauthorized collection and processing of such personal data can lead to legal problems. There are data fiduciary duties, notice, and individual rights on access, correction, and erasure in the Act [11].

Though the DPDP Act was enacted to regulate the collection and processing of personal data, it adds to the legal weight that consent, misuse, and platforms have to consider in cases involving personal data processed without authorization.

E. Bharatiya Nyaya Sanhita, 2023

The Bharatiya Nyaya Sanhita, 2023 (BNS) reforms India's laws on criminal matters [12]. Though it has not yet provided a specific deepfake offence in the criminal code, offences like cheating, deception, obscenity, harassment, defamation, and acts against women might apply according to circumstances and intention. Indeed, discussions in India have been advocating for the addition of more deepfake-related penal laws under the BNS legislation [22].

F. Gaps in the Indian Framework

However, in spite of the presence of several legal measures applicable to various forms of harm caused by the use of deepfakes, there remain numerous deficiencies in India's regulatory and enforcement measures concerning the problem of misuse of synthetic media. Specifically, as of now, the Indian legal framework does not provide for any statutory definition of the term "deepfake" or legal classification for the use of synthetic media causing harm to the user's well-being. Furthermore, there is a notable absence of a legal mechanism aimed at the protection of individuals' right to their image, voice, and digital identity. The problem is further exacerbated by the insufficient legal means for establishing proof concerning the use of synthetic media in the course of an ongoing dispute. Finally, as far as enforcement is concerned, it is often contingent upon the voluntary compliance of the social media platform in question and reporting by the potential victim. This means that cross-border enforcement of rules and regulations in the realm in question is quite challenging due to the global nature of the internet and relatively slow pace of legal measures as compared to the speed of circulation of abusive content.

VII. GLOBAL REGULATORY DEVELOPMENTS

While this paper highlights India's case, trends on the international stage are instructive.

One of the important legal instruments aimed at addressing the issue of synthetic media is the EU AI Act. Available public summaries and legal analysis suggest that the act will impose transparency requirements obliging people to notify whenever images, audio, or video content have been artificially generated or edited in ways that can make them appear real, except for some narrow exceptions [23], [24]. In other words, the governance system will be built around labeling, transparency, and accountability rather than solely relying on criminalization.

Furthermore, there is a trend toward increased concern expressed by European officials in charge of regulation and privacy about the abuse of deepfake technology in connection with sexually explicit images [25]. This implies that there may be an increased effort to include some aspects of AI law, privacy law, platform governance, and cybersecurity law in regulatory regimes.

India can certainly do the same, although with some modifications based on its unique circumstances.

VIII. CHALLENGES IN FUSION OF DETECTION AND LAW

A. Technical Uncertainty As Legal Uncertainty

Another issue is that deepfake detection technology is not foolproof.If there are false positives, legitimate material might be incorrectly marked. If there are false negatives, dangerous material will still be online. During litigation, there will be issues regarding admissibility, evidentiary value, and proof.

B. The Liar's Dividend

With synthetic media proliferating, perpetrators could argue that legitimate evidence is manipulated. This affects confidence in authentic materials and complicates criminal investigations, employment disputes, and governmental accountability [18].

C. Jurisdictional and Platform Issues

The offender is located in one country, the storehouse in another country, the victim in a third country, and the recipients in all countries. Therefore, it becomes difficult to apprehend, track, and prosecute the crime.

D. Speed of Harm versus Speed of Law

A deepfake video spreads rapidly. However, legal notice, police report, forensic analysis, and judicial remedy will all take far longer than that. At the point when a legal response is implemented, the harm to society might have already become irreversible.

E. The Requirement for Human Interference

Guilt, censorship, or deletion of the material must never be automated in cases of this sensitivity. Human interference is essential, especially where satire, parody, journalism, art, or politics is involved.

IX. RECOMMENDATIONS AND FUTURE DIRECTIONS

In tackling the problem of deepfakes, a comprehensive approach that considers both legal and technical solutions would be required. As an example, there is a need for India to develop laws regarding deepfakes, including the definition of deepfakes, creative/artistic deepfakes, impersonating/deceptive deepfakes, and malicious/fraudulent deepfakes. Moreover, a harm-based regulatory regime can be used to determine varying responses to different kinds of synthetic content based on their degree of harm and impact on users. This regime would recognize the difference between harmless or labeled content, misleading or privacy-invasive content, fraudulent or sexually exploitative deepfakes, and those which may pose electoral and public threats. Another important requirement is greater intermediary liability whereby digital platforms need to develop procedures to provide rapid redressal for complaints against fake news, retain forensic metadata related to any detected synthetic media, label the content with a visible indicator, ensure traceability through audit logs, and have an escalation process for risky content according to existing due diligence requirements in India [10], [13], [14]. Lastly, national forensic standards for deepfakes in terms of chain of custody, authentication process, confidence rating, testimony, admissibility of evidence, and model documents need to be developed. The significance of forensic methods that use this type of systematic framework can be emphasized by the latest findings by NIST, which point to the difficulties associated with assessing AI-produced evidence and the necessity for effective methodologies [19].

X. CONCLUSION

Among the most apparent cases of how AI can present challenges to technology as well as to law at the same time are deepfakes. On one hand, such media present a very blurred boundary between reality and forgery. On the other hand, they demonstrate the limitations of the traditional legal system that was never designed to regulate cases involving artificial manipulations with one's digital identity. As deepfaking algorithms become easier to use and more realistic, their impact goes far beyond potential violations of copyright and artistic creation.

This paper has provided evidence of how detection of deepfakes advances due to traditional forensic analysis techniques, deep learning-based algorithms, multimodal approaches, and explainable AI. Nevertheless, there are several aspects to consider in order to acknowledge the fact that existing solutions are still imperfect and may lack accuracy in real-world conditions when deepfakes get generated differently and compressed or modified.

With respect to cyberspace, India already uses a combination of legislation that includes the IT Act, IT Rules, DPDP Act, and common criminal law. This legislation can help deal with deepfake problems indirectly; however, no specialized approach exists at the moment. In the lack of legal definitions, remedies, forensic protocols, and prompt regulatory measures, there is an obvious regulatory lacuna.

Ultimately, though, whatever will be required to manage the phenomenon of deepfakes cannot be accomplished exclusively either technologically or legally. An ecological approach that includes better detection, forensic transparency, platform responsibility, privacy protection, media literacy, and innovative legislation will be required. Societies capable of preserving the integrity of digital evidence in combination with freedom of innovation and expression will find themselves well-prepared for the deepfake challenge.

REFERENCES

[1] B. C. Soundarya and H. L. Gururaj, “Deepfake detection: critical review of state-of-the-art approaches and future perspectives,” Discover Applied Sciences, vol. 8, art. no. 201, Jan. 2026. (Springer)

[2] A. A. Khan, A. A. Laghari, S. A. Inam, S. Ullah, and M. Shahzad, “A survey on multimedia-enabled deepfake detection: state-of-the-art tools and techniques, emerging trends, current challenges & limitations, and future directions,” Discover Computing, vol. 28, art. no. 48, Apr. 2025. (Springer)

[3] “Deepfake fraud taking place on an industrial scale, study finds,” The Guardian, Feb. 6, 2026. (The Guardian)

[4] A. M. Panagopoulos and A. Davalas, “Deepfakes on the EU AI Act and Its Implementation in the Newsrooms,” International Journal of Social Science and Economic Research, vol. 10, no. 8, pp. 3276–3296, Aug. 2025. (ResearchGate)

[5] A. Raza, A. Basit, A. Amin, Z. A. Arfeen, M. I. Masud, U. Fayyaz, and T. A. Jumani, “A Comprehensive Review of Deepfake Detection Techniques: From Traditional Machine

Learning to Advanced Deep Learning Architectures," AI, vol. 7, no. 2, art. no. 68, Feb. 2026. (MDPI)

[6] D. Tan, Y. Yang, C. Niu, S. Li, D. Yang, and B. Tan, "A review of deep learning based multimodal forgery detection for video and audio," Discover Applied Sciences, vol. 7, art. no. 987, Aug. 2025. (Springer)

[7] D.-A. Boldisor, S. Smeu, D. Oneata, and E. Oneata, "Investigating self-supervised representations for audio-visual deepfake detection," arXiv preprint arXiv:2511.17181, 2025. (arXiv)

[8] S. Tariq, S. S. Woo, P. Singh, I. Irmalasari, S. Gupta, and D. Gupta, "From Prediction to Explanation: Multimodal, Explainable, and Interactive Deepfake Detection Framework for Non-Expert Users," arXiv preprint arXiv:2508.07596, 2025. (arXiv)

[9] Government of India, The Information Technology Act, 2000, Act No. 21 of 2000. New Delhi, India: Ministry of Electronics and Information Technology, 2000. (India Code)

[10] Government of India, The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. New Delhi, India, 2021.

[11] Government of India, The Digital Personal Data Protection Act, 2023, Act No. 22 of 2023. New Delhi, India: Ministry of Electronics and Information Technology, 2023. (India Code)
[12] Government of India, The Bharatiya Nyaya Sanhita, 2023. New Delhi, India, 2023. (India Code)

[13] Press Information Bureau, Ministry of Electronics & Information Technology, "MeitY issues advisory to all intermediaries to comply with existing IT rules," Dec. 26, 2023. (Press Information Bureau)

[14] Press Information Bureau, Ministry of Electronics & Information Technology, "Union Government issues advisory to social media intermediaries to identify misinformation and deepfakes," Nov. 7, 2023. (Press Information Bureau)

[15] Y. Li and S. Lyu, "Exposing DeepFake Videos By Detecting Face Warping Artifacts," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), 2019.

[16] H. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule-Forensics: Using Capsule Networks to Detect Forged Images and Videos," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2019.

[17] R. Chesney and D. K. Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," California Law Review, vol. 107, no. 6, pp. 1753–1820, 2019.

[18] B. Vaccari and A. Chadwick, "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust," Social Media + Society, vol. 6, no. 1, 2020.

[19] H. Guan, J. Horan, and A. Zhang, "Guardians of Forensic Evidence: Evaluating Analytic Systems Against AI-Generated Deepfakes," NIST, Forensics@NIST 2024, published Jan. 27, 2025. (NIST)

[20] K. Gandhi, P. Kulkarni, T. Shah, P. Chaudhari, M. Narvekar, and K. Ghag, "A Multimodal Framework for Deepfake Detection," arXiv preprint arXiv:2410.03487, 2024. (arXiv)

[21] Y. Du, Z. Wang, Y. Luo, C. Piao, Z. Yan, H. Li, and L. Yuan, "CAD: A General Multimodal Framework for Video Deepfake Detection via Cross-Modal Alignment and Distillation," arXiv preprint arXiv:2505.15233, 2025. (arXiv)

[22] "NCW seeks penalties under BNS to counter deep fake abuse," The Times of India, Nov. 2025. (The Times of India)

[23] European Parliament, "Artificial Intelligence Act: MEPs adopt landmark law," Press Release, Mar. 2024.

[24] A. M. Panagopoulos and A. Davalas, "Deepfakes on the EU AI Act and Its Implementation in the Newsrooms," International Journal of Social Science and Economic Research, vol. 10, no. 8, pp. 3276–3296, 2025. (ResearchGate)

[25] "Italy's privacy watchdog warns Grok over deepfake AI content," Reuters, Jan. 8, 2026. (Reuters)