



RESEARCH ARTICLE

A Comprehensive Study on Addressing Trust Erosion in Multimedia in The Indian Context

Kshema Manu¹ and Malathi S^{2*}

Abstract

The widespread use of multimedia helps users to connect, create, and inform easily, but the consequences of inappropriate usage, including altered content, deepfakes, and misleading material, are severely damaging. It can undermine trust, influence public opinion, deceive people, go against Government rules and regulations, and even incite violence due to its inherent characteristic of being quick and widespread. Spreading illicit material, committing cyberbullying, and harassing individuals are a few cases of multimedia misuse. An integrated approach combining technology, digital literacy, critical thinking, guidelines, and timely legal reforms is the only possible solution to curb the complexity of multimedia misuse. The observations made in this paper advocate for stronger legislative measures and cross-disciplinary collaboration to address the evolving landscape of trust erosion.

Keywords: Algorithms, BNS, Deepfake, Disinformation, IPC, IT Act, Journalism, Law, Misinformation, Multimedia, Trust.

Introduction

Multimedia encompasses text, images, audio, video, graphics, animations, AR/VR, and more (McGaughey, 2001). It effectively captures user attention and reaches a wide audience. Bob Goldstein coined the term "Multimedia" in 1966 for his performance art event "Lightworks" (Meyers, 2000). Multimedia faces challenges like plagiarism, content misuse, harmful material, data distortion, and privacy violations. Trust is eroded by disinformation, copyright issues, technical limitations, and quality degradation (Dwivedi & Pachauri, 2023). Social issues like the digital divide, cultural exploitation, and content overload desensitize audiences (Youvan, 2024). Emerging threats like deepfakes

and AI-generated content further undermine trust (Lyu, 2024). Understanding these challenges and their solutions is crucial.

Materials and Methods

This section outlines the methodology employed for the preparation and conduct of the research presented in this paper.

Keyword Selection and Resource Accumulation

Relevant keywords were identified to guide the collection of material. Key terms such as "AI-generated content," "deepfake impact," "deepfake detection," "trust in synthetic media," "deepfake prevention," and "multimedia authentication" were used to search Google Scholar, the primary database for content extraction. To ensure the rigor and validity of the research, inclusion criteria were established: articles had to be peer-reviewed journal or conference papers, government reports, or government regulations, and published between 2000 and 2024. The distribution of included articles is depicted in Figure 1.

Content Analysis

Applying the inclusion criteria, 241 articles were initially selected. After careful examination, 116 articles and 24 legal instruments were ultimately chosen for organization, analysis, and review. Key information was extracted from each article, facilitating a comparative analysis. The distribution of reference materials across different concerns, as shown in Figure 2, indicates the primary research areas

¹Department of Computer Applications, Cochin University of Science and Technology, Kochi - 682022

²Department of Computer Applications, Cochin University of Science and Technology, Kochi - 682022

***Corresponding Author:** Malathi S, Department of Computer Applications, Cochin University of Science and Technology, Kochi - 682022, E-Mail: malathi_s@cusat.ac.in

How to cite this article: Manu, K., Malathi S. (2025). A Comprehensive Study on Addressing Trust Erosion in Multimedia in The Indian Context. *The Scientific Temper*, 16(3):3887-3896.

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.3.05

Source of support: Nil

Conflict of interest: None.

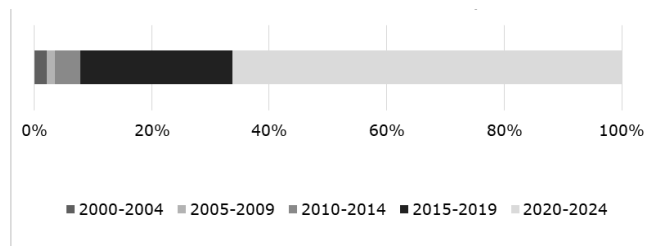


Figure 1: Reference Distribution by year

within the domain. A higher number of references correlates with increased research activity in a specific area.

Information Obtained on Various Concerns

Issues Arising from Misuse of Multimedia

The misuse of multimedia technologies, including AR/VR and biometric data collection, raises significant ethical and societal concerns. These include unauthorized data collection, privacy violations, and the potential for addiction. “Always-on” devices inadvertently gather sensitive information, necessitating safeguards like adjustable immersion levels (García et al., 2020; Kade, 2015; Royakkers et al., 2018; Shahbodin et al., 2024; Slater et al., 2020).

The rise of deepfakes and misinformation undermines public trust in digital platforms, emphasizing the need for accuracy in journalism and ethical content practices (Shirky, 2014). Cultural insensitivity in content creation risks stereotypes, underscoring the importance of ethical guidelines (Shahbodin et al., 2024). The rise of AI-powered tools, such as deepfakes and chatbots, increases the risk of identity theft and fraud (Irshad & Soomro, 2018).

The digital realm can intensify existing societal challenges. For example, cyberbullying disproportionately affects marginalized groups like women, individuals from the LGBTQIA+ community, etc. (Maji & Abhiram, 2023) and excessive screen time driven by engagement algorithms negatively impacts mental health as it reduces face-to-face interactions (Karim et al., 2020; Shabahang et al., 2022).

Media misinformation and biased reporting further erode public trust, deepening polarization and undermining the media’s role as the “fourth estate” (Michailidou & Trenz, 2021).

Factors Contributing to Misuse

Disinformation/Misinformation

Misinformation is spread through manipulated images and videos, which is facilitated by automated chatbots and AI-generated or unverified content (Anthonysamy and Sivakumar, 2022; Dufour et al., 2024). The use of emotionally charged content further increases the sharing rate of such content and thus aids polarization (Wan et al., 2024). Since 2023, there has been a notable rise in the prevalence of “echo chambers,” small groups that amplify misinformation within their closed networks (Dufour et al., 2024).

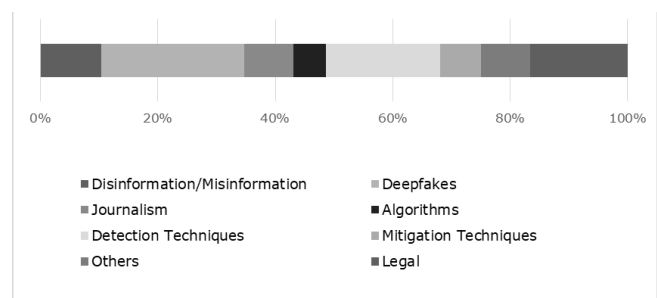


Figure 2: Reference Distribution by Subject Area

Deepfakes

Deepfakes are fake media that spread false information and impersonate individuals, thereby undermining trust in multimedia platforms (Millière, 2022). Generative Adversarial Networks (GANs) are neural networks that generate new media through a competitive process between a generator and a discriminator (Masood et al., 2022). Popular GAN architectures are DCGAN (Radford et al., 2015), WGAN (Arjovsky et al., 2017), PGGAN (Karras et al., 2017), and StyleGAN (Karras et al., 2019). Diffusion models create images by adding noise to a clear image and then using a neural network to remove it step-by-step (Lyu, 2024). Since 2021, systems like Midjourney and Stable Diffusion have significantly improved image generation (Kingma & Welling, 2014). Large Language Models (LLMs) use transformers to generate human-like text (Lyu, 2024), while Variational Autoencoders (VAEs) can swap faces while preserving unique facial features. These technologies have significantly improved image generation since 2021 (Kingma & Welling, 2014).

Declining Journalism Standards

Sensationalism, clickbait, commercialization, and the blurred lines between news and opinion erode journalistic standards (Oyinloye et al., 2024; Youvan, 2024). Sensational headlines and corporate influence prioritize profits over accuracy and investigative journalism (Farid, 2023; Pepple & Acholonu, 2018). Lack of fact-checking spreads unverified information, while algorithmic journalism risks bias and disinformation (Dörr & Hollnbuchner, 2016; Jastaniyah, 2017; Martyn, 2009; Pavlik, 2023; Thomson et al., 2020). Ethical decline due to poor compensation and job insecurity further exacerbates the issue (Herzog, 2021; Ruggiero et al., 2022).

Algorithmic Bias

Social media algorithms, such as those used by Facebook and YouTube, can influence political polarization and public trust in media. Personalization algorithms, which customize content based on user behavior, can narrow information scope (Kitchens et al., 2020). Recommendation systems, such as collaborative filtering and content-based filtering, can promote engaging content at the cost of diversity.

Collaborative filtering can recommend content based on user similarity (Sarwar et al., 2001), while content-based filtering suggests content similar to past preferences (Fayyaz et al., 2020). Engagement-based algorithms prioritize content that generates high engagement, often favoring sensational or polarizing content (Kitchens et al., 2020). For example, Facebook’s promotion of partisan content, which garners more engagement, can deepen ideological divides (Bakshy et al., 2015).

Existing Legal Frameworks in India

India’s legal framework is complex, incorporating existing laws and regulatory guidelines to address technology, free expression, and societal safeguards. The Indian Evidence Act (IEA)¹, 1862, outlines the admissibility of electronic evidence in court. Amendments to Section 65B have broadened the definition of electronic records and relaxed admissibility requirements in Section 63 of the Bharatiya Sakshya Adhiniyam (BSA)², 2023. The amendments include semiconductor memory devices, expanded storage methods, and a focus on reliability and authenticity. The person in charge of the device or relevant activities can now issue certificate requirements for electronic records.

The *Ajit Mohan*³ case addressed the conflict between the right of corporations to privacy and the government’s responsibility to oversee digital activities and the Supreme Court affirmed the authority of the legislative committee. while in the *Vitus*⁴ case the top court emphasized individual privacy rights, finding excessive bail conditions unconstitutional.

The *Hajam*⁵ and *Anu Kumari*⁶ Cases explored the legal implications of social media posts, particularly WhatsApp statuses. The Supreme Court underscored the importance of free speech, ruling that non-hateful posts don’t warrant criminal charges. In Anu Kumari, the court highlighted the seriousness of the allegations and denied anticipatory bail.

India addresses multimedia misuse through a multi-faceted legal framework, though gaps remain in tackling emerging challenges like deepfakes and algorithmic manipulation. Relevant legislation is outlined in Table 1.

Table 1: Legal Landscape Addressing Misuse of Multimedia in India

Legislation	Relevant Sections	Key Provisions
Indian Penal Code (IPC) ⁷ , 1860 and Bharatiya Nyaya Sanhita (BNS) ⁸ , 2023	IPC Sections: 153A, 153B, 295A, 416, 500, 505, 177, 157; BNS Sections: 196, 197, 299, 356, 353, 201, 212	Promotion of enmity, defamation, prejudicing national integration, outraging religious sentiments, cheating, mischief, misinformation
Information Technology (IT) Act ⁹ , 2000	Sections 66, 67, 69A, 79	Computer-related offenses, Publishing/transmitting obscene material, blocking access to online content (66A struck down ¹⁰)
Protection of Children from Sexual Offences (POCSO) Act ¹¹ , 2012	Sections 13, 14, 15	Child pornography, storage of such material
Representation of the People Act ¹² , 1951	Sections 123, 125	Bribery, undue influence, intimidation, false statements in election manifestos, class enmity.
Indian Copyright Act ¹³ , 1957	Sections 13, 14, 15, 62, 63	Copyright protection, infringement penalties
Constitution of India ¹⁴	Article 21	Right to Privacy
Trade Marks Act ¹⁵ , 1999	Sections 29, 52	Infringement of trademarks, punishments

1 The Indian Evidence Act, 1862 (IEA). (1860). Ministry of Law and Justice, Government of India, Act No. 1 of 1872

2 Bharatiya Sakshya Adhiniyam, 2023 (BSA). (2023). Ministry of Law and Justice, Government of India, Act No. 47 of 2023

3 Ajit Mohan & Ors. v. Legislative Assembly, National Capital Territory of Delhi & Ors., Writ Petition (C) No. 1088 of 2020, Supreme Court of India, 2020

4 Vitus, F. v. Narcotics Control Bureau & Ors., Criminal Appeal No. of 2024, Supreme Court of India, 2024.

5 Hajam, J. A. v. State of Maharashtra & Anr., Criminal Appeal No. 886 of 2024, Supreme Court of India, 2024

6 Anu Kumari v. State of Punjab & Ors., Civil Revision No. 721 of 2004 (O&M), High Court of Punjab and Haryana, 2012.

7 The Indian Penal Code, 1860 (IPC). (1860). Ministry of Law and Justice, Government of India, Act No. 45 of 1860.

8 Bharatiya Nyaya Sanhita, 2023. (2023). Ministry of Law and Justice, Government of India, Act No. 45 of 2023

9 Information Technology Act, 2000. (2000). Ministry of Electronics and Information Technology, Government of India

10 Shreya Singhal v. Union of India, Writ Petition (Criminal) No. 167 of 2012, Supreme Court of India, 2015.

11 Protection of Children from Sexual Offences Act, 2012. (2012). Ministry of Women and Child Development, Government of India

12 Representation of the People Act, 1951. (1951). Ministry of Law and Justice, Government of India, Act No. 43 of 1951

13 The Copyright Act, 1957, Section 14. (1957). Ministry of Commerce and Industry, Government of India

14 India. (1950). The Constitution of India. [Dehra Dun: Photolithographed at the Survey of India Offices, 195] [Pdf] Retrieved from the Library of Congress, <https://www.loc.gov/item/57026883/>

15 India. (1999). The Trade Marks Act, 1999. [Act No. 47 of 1999.] Retrieved from India Code: <https://www.indiacode.nic.in/bitstream/123456789/1993/1/A1999-47.pdf>

The Medical Council Code of Ethics Act¹⁶, SEBI Act¹⁷, Environmental Protection Act¹⁸, Consumer Protection Act¹⁹, Bar Council of India Regulations²⁰, Chartered Accountants Act²¹, Companies Act²², Press Information Bureau (PIB) Fact Check Unit, and Press Council of India (PCI) are all laws aimed at promoting ethical practices in healthcare, financial market transparency, environmental protection, consumer protection, professional integrity, corporate governance, and journalistic ethics.

The Digital Personal Data Protection Act²³, 2023, and the Information Technology Act Rules²⁴, 2011, address privacy and consumer protection. The Consumer Protection Act, 2019, and the Environment Protection Act also have relevant provisions. Social media platforms' data practices may be indirectly regulated by the Digital Personal Data Protection Act, 2023. The Digital India Act addresses concerns like algorithmic bias, internet privacy, and content regulation. Understanding these laws can help individuals and organizations avoid legal repercussions and ensure responsible information sharing.

Existing laws provide a foundation, but challenges remain. The lack of specific regulations for deepfakes and limited algorithmic accountability hinders progress. International cooperation is needed to address cross-border cybercrime. A comprehensive approach, including legislative reforms and digital literacy, is crucial to restore trust in the digital realm.

Technological Approaches for Detection

To combat misinformation and deepfakes, various techniques are employed:

Text Analysis

Transformer-based models like BERT and RoBERTa capture word context and semantics for text-based tasks (Ayetiran & Özgöbek, 2024).

Image Analysis

Convolutional Neural Networks (CNNs) like ResNet are used for image analysis to learn complex features and detect image manipulations (Abukari et al., 2023; Ghai et al., 2021).

Audio Analysis

WaveNet and MFCC+CNN are used for audio analysis, while LSTMs capture temporal dependencies (Ayetiran & Özgöbek, 2024).

Video Analysis

I3D and X3D extend 2D CNNs into 3D for video analysis (Hu et al., 2021b). LSTM-CNN Hybrid models combine CNNs and LSTMs for enhanced video-based fake news detection (Ayetiran & Özgöbek, 2024).

- Multi-modal analysis models like EANN and MVAE combine text and image features, factoring in event-specific biases using variational autoencoders (Ayetiran & Özgöbek, 2024). SpotFake and SAFE measure similarities between text and images (Ayetiran & Özgöbek, 2024). The Emotion-Aided Multi-task Framework uses emotional analysis from text, audio, and video to boost accuracy, especially for emotionally charged content (Kumari et al., 2023).
- Knowledge-Enhanced Models: VisualBERT incorporates external knowledge from large-scale knowledge graphs for improved fake news detection (Gao et al., 2024a).
- Pixel-Level Analysis: Techniques such as Pixel-Level Disparities (Li et al., 2020), Saturated Pixel Analysis (McCloskey & Albright, 2019), Co-occurrence Matrices (Nataraj et al., 2019), Photo Response Non-Uniformity (PRNU) Pattern (Koopman et al., 2018), Hybrid Spatial and Frequency Analysis (Liu et al., 2021; Masi et al., 2020), Local Motion Features (Wang et al., 2020b), Corneal Specular Highlight (Hu et al., 2021a), and Frequency Spectrum Analysis (Barni et al., 2020; Frank et al., 2020) can be used to analyze individual pixels.
- Artifact Analysis: Techniques like Local Patch Artifacts (Chai et al., 2020), Artifact-Based Detection (Nirkin et al., 2020), and Texture and Artifact Detection (TAD) can be used to identify specific artifacts or inconsistencies in images.
- Traditional machine learning models, such as Support Vector Machines (SVM), Random Forests, and Naive Bayes (Malik et al., 2022), classify deepfakes based on specific attributes.
- Deep learning methods: CNNs extract features from images or videos to classify them as real or fake (Guarnera et al., 2020; Guo et al., 2020; Wang et al., 2020a). While recurrent neural networks (RNN) analyze temporal dependencies in video sequences (de Lima et al., 2020; Sabir et al., 2019). GANs generate synthetic data for training or analyzing the structure of deepfakes (Marra et al., 2019; Zhang et al., 2019).

16 Indian Medical Council Code of Ethics. (n.d.). Medical Council of India

17 Securities and Exchange Board of India (SEBI) Act, 1992. (1992). Ministry of Corporate Affairs, Government of India

18 Environment Protection Act, 1986. (1986). Ministry of Environment, Forest and Climate Change, Government of India

19 Consumer Protection Act, 2019. (2019). Ministry of Consumer Affairs, Food & Public Distribution, Government of India

20 Bar Council of India Regulations. (n.d.). Bar Council of India

21 The Chartered Accountants Act, 1949. (1949). Ministry of Corporate Affairs, Government of India

22 The Companies Act, 2013. (2013). Ministry of Corporate Affairs, Government of India

23 Digital Personal Data Protection Act, 2023. (2023, August 11). Gazette of India, 22 of 2023.

24 Government of India. (2011). The Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011. Gazette of India.

Other deep learning techniques include Cross Layer Refinement Network (CLRNet) (Tariq et al., 2020), Sliced Spatio-Temporal Network (SSTNet) (Wu et al., 2020), Hierarchical Memory Network (HMN) (Fernando et al., 2019), Gram-Net (Liu et al., 2020), Fake Detection Fine-tuning Network (FDFtNet) (Jeon et al., 2020), FakeLocator (Huang et al., 2022), Manipulation Tracing Network (ManTra-Net) (Wu et al., 2019), OC-FakeDect (Khalid & Woo, 2020), Siamese Network with Triplet Loss (Agarwal et al., 2020a; Mittal et al., 2020), Multi-task Learning (Nguyen et al., 2019), Attention Map Estimation (Dang et al., 2020), and Octave Convolution (OctConv) (Chen et al., 2019; Yu et al., 2020).

- **Biological and Behavioural Analysis:** These techniques use physiological and behavioral cues to detect deepfakes. Facial analysis including lip-sync inconsistencies (Korshunov & Marcel, 2018), Layer-by-Layer Neuron Behaviour (Wang et al., 2019), eye blinking (Li et al., 2018), Deep Face Recognition Systems (Nhu et al., 2018), and head pose and landmark analysis (Yang et al., 2019a; Yang et al., 2019b), and physiological signal analysis, such as heart rate analysis (DeepRhythm and MMSTR) (Li & Lyu, 2019; Qi et al., 2020), remote photoplethysmography (rPPG) (Ciftci et al., 2020a; Ciftci et al., 2020b), Heart Rate Detection (Fernandes et al., 2019; Qi et al., 2020), and skin color changes due to heartbeat (Hernandez-Ortega et al., 2020; Qi et al., 2020), are useful in identifying deepfakes.
- **Contextual and feature-based techniques** focus on the context and specific features of the image/video. Techniques like Face-Context Discrepancy (Nirkin et al., 2020), Facial and Head Movement Patterns (Agarwal et al., 2019; Agarwal et al., 2020b), Facial Key Point Detection (Zhang et al., 2017), Fixed Size Faces Detection (Li & Lyu, 2019), and Modality Dissonance Score (MDS) (Chugh et al., 2020) are used to detect anomalies. Additionally, techniques like Localization of Manipulated Regions (Huang et al., 2022) and variations in face sizes (Coccomini et al., 2024) can also be employed.
- **Preprocessing and Quality Measures:** Techniques like Dataset Preprocessing Techniques (Chen & Yang, 2021), Image Preprocessing Techniques (Xuan et al., 2019), No-Reference (NR), Full-Reference (FR) Quality Measures (Concas et al., 2024), and Impact of Compression and Detection under Compression (Galvan et al., 2013; Gao et al., 2024b) are used to prepare and process data to enhance detection accuracy.
- **Forensic and Device-based Approaches:** Forensic Localization Dataset (Songsri-in & Zafeiriou, 2019) and DCT Coefficient Analysis (Battiato & Messina, 2009) are used to identify manipulated media. Analyzing facial features, patterns in audio and video, inconsistencies in lighting, shading, and other visual cues like the absence

of background noise or echoes can also be helpful. Device-based signatures such as Media Signature Encoding (Baracchi et al., 2024) and Watermarking (Yu et al., 2021) are embedded into media files to track their origin and detect tampering. GAN Watermarking (Fei et al., 2022) is a technique that uses GANs to embed robust watermarks.

- **Transferability analysis** studies the ability of deepfake detection models to generalize across different datasets and domains (Barni et al., 2019)
- While these approaches show promise, the rapid evolution of generative technologies challenges their effectiveness (Lyu, 2024).
- *Strategic Mitigation Methods*
 - Enhance algorithmic transparency in personalization and recommendation systems (Pariser, 2012).
 - Prioritize Digital Literacy Education among users to recognize misinformation, cyber-crime, and deepfakes (Wan et al., 2024).
- *Strengthen Regulatory Frameworks*
 - Foster collaborative fact-checking initiatives, like Full Fact, use hybrid systems combining human oversight and ML for real-time verification (Nekmat, 2020). Collaboration between fact-checkers and media platforms can address misinformation (Graves, 2018).
 - Enhance platform accountability by demanding strict content moderation policies (Anwar & Fong, 2012; Lazer et al., 2018).
- *Restoring journalistic integrity*
 - Technological and cognitive literacy, along with social-emotional skills, can help reduce misinformation's impact (Gaillard et al., 2021; Pal et al., 2019).

To effectively mitigate the misuse of multimedia, a comprehensive approach, combining these strategies and adapting to evolving technologies, is necessary.

Discussion

The erosion of trust in multimedia is driven by deepfakes, misinformation/disinformation, algorithmic bias, and declining journalistic standards, which align with prior research. Deepfakes are powered by GANs and diffusion models, which pose significant threats (Arjovsky et al., 2017; Radford et al., 2015). While transformer-based models show promise in their effectiveness in detecting deepfakes, it is often limited by the adaptability of generative models (Li et al., 2020; Nataraj et al., 2019). Algorithms foster echo chambers and filter bubbles, which amplify misinformation and political polarization (Bakshy et al., 2015; Kitchens et al., 2020). Hybrid systems are one example of advanced detection methods that can offer potential solutions (Hassan et al., 2017; Nekmat, 2020). Commercialization and sensationalism are major contributors to declining

journalistic standards, and they proportionally erode public trust (Oyinloye et al., 2024; Pepple and Acholonu, 2018). While fact-checking and effective training can help rebuild trust (Cavaliere, 2020; Graves, 2018), India's legal framework needs adaptation to address challenges such as deepfakes and algorithmic bias.

Thus a multi-faceted approach that has legislative reforms, technological advancements in detecting harmful content, and media literacy is essential for restoring trust in multimedia.

Conclusion

The corrosive nature of false information erodes trust, distorts public opinion, and thwarts decision-making. Major contributing factors are deepfakes, declining journalistic standards driven by sensationalism and commercialization, and algorithmic bias, which amplify misinformation and polarization.

A systematic approach combining legislation, education, international cooperation, media literacy, and advancements in detection technology is crucial. Independent reporting, increasing accountability, and ethical journalism are also essential.

Future research should prioritize algorithms that promote diversity, user control, transparency, and advanced detection techniques, including deep learning and hybrid models, which can combat fake news and privacy issues. Ethical considerations are paramount in multimedia practices, and therefore, professionals must uphold the highest standards to ensure a credible and reliable digital environment.

Future Scope

While progress has been made in detecting and mitigating misinformation and deepfakes, challenges remain. Real-time detection, cross-modal analysis, ethical considerations, user behavior, and international cooperation require further exploration. Future research should focus on AI's impact on journalism standards, sustainable business models for ethical journalism, audience trust metrics, and leveraging blockchain, IoT, and deep learning for improved recommendation systems and fake content detection. Addressing the misuse of multimedia necessitates a comprehensive approach. This includes strengthening legal frameworks, promoting digital literacy, and fostering ethical practices. International cooperation is crucial to combat cross-border issues like deepfakes and misinformation.

Acknowledgement

We, the authors of this work, have not received support or help from any other persons or professionals. We didn't have any collaborators, either. This work we have conducted is purely academic research. We owe thanks to the Almighty and our family, friends, and colleagues for their emotional support.

Conflict Of Interest

The authors declare no conflict of interest. This study's work and findings reflect the authors' independent research and analyses without any interference, collaboration, or influence from other persons or organizations.

References

- Abukari, A. M., Madavarapu, J. B., & Bankas, E. K. (2024). A lightweight algorithm for detecting fake multimedia contents on social media. *Earthline Journal of Mathematical Sciences*, 14(1), 119-132.
- Agarwal, S., Farid, H., El-Gaaly, T., & Lim, S. N. (2020a, December). Detecting deep-fake videos from appearance and behavior. In *2020 IEEE international workshop on information forensics and security (WIFS)* (pp. 1-6). IEEE.
- Agarwal, S., Farid, H., Fried, O., & Agrawala, M. (2020b). Detecting deep-fake videos from phoneme-viseme mismatches. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 660-661).
- Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., & Li, H. (2019, June). Protecting world leaders against deep fakes. In *CVPR workshops* (Vol. 1, No. 38).
- Anthonyamy, L., & Sivakumar, P. (2022). A new digital literacy framework to mitigate misinformation in social media infodemic. *Global Knowledge, Memory and Communication*, 73(6/7), 809-827.
- Anwar, M., & Fong, P. W. (2012, March). A visualization tool for evaluating access control policies in facebook-style social network systems. In *Proceedings of the 27th annual ACM Symposium on Applied Computing* (pp. 1443-1450).
- Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein gan. *arXiv preprint arXiv: 170107875*. arXiv preprint arXiv:1701.07875.
- Ayetiran, E. F., & Özgöbek, Ö. (2024). A review of deep learning techniques for multimodal fake news and harmful languages detection. *IEEE Access*.
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130-1132.
- Baracchi, D., Boato, G., De Natale, F., Iuliani, M., Montibeller, A., Pasquini, C., ... & Shullani, D. (2024). Towards open-world multimedia forensics through media signature encoding. *IEEE Access*.
- Barni, M., Kallas, K., Nowroozi, E., & Tondi, B. (2019, May). On the transferability of adversarial examples against CNN-based image forensics. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 8286-8290). IEEE.
- Barni, M., Kallas, K., Nowroozi, E., & Tondi, B. (2020, December). CNN detection of GAN-generated face images based on cross-band co-occurrences analysis. In *2020 IEEE international workshop on information forensics and security (WIFS)* (pp. 1-6). IEEE.
- Battiato, S., & Messina, G. (2009, October). Digital forgery estimation into DCT domain: a critical analysis. In *Proceedings of the First ACM workshop on Multimedia in forensics* (pp. 37-42).
- Cavaliere, P. (2020). From journalistic ethics to fact-checking practices: defining the standards of content governance in the fight against disinformation. *Journal of media law*, 12(2), 133-165.

- Chai, L., Bau, D., Lim, S. N., & Isola, P. (2020). What makes fake images detectable? understanding properties that generalize. In *Computer vision—ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, part XXVI* 16 (pp. 103-120). Springer International Publishing.
- Chen, Z., & Yang, H. (2021, June). Attentive semantic exploring for manipulated face detection. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1985-1989). IEEE.
- Chen, Z., Tondi, B., Li, X., Ni, R., Zhao, Y., & Barni, M. (2019). Secure detection of image manipulation by means of random feature selection. *IEEE Transactions on Information Forensics and Security*, 14(9), 2454-2469.
- Chugh, K., Gupta, P., Dhall, A., & Subramanian, R. (2020, October). Not made for each other-audio-visual dissonance-based deepfake detection and localization. In *Proceedings of the 28th ACM international conference on multimedia* (pp. 439-447).
- Ciftci, U. A., Demir, I., & Yin, L. (2020a). Fakecatcher: Detection of synthetic portrait videos using biological signals. *IEEE transactions on pattern analysis and machine intelligence*.
- Ciftci, U. A., Demir, I., & Yin, L. (2020b, September). How do the hearts of deep fakes beat? Deep fake source detection via interpreting residuals with biological signals. In *2020 IEEE international joint conference on biometrics (IJCBI)* (pp. 1-10). IEEE.
- Coccomini, D. A., Zilos, G. K., Amato, G., Caldelli, R., Falchi, F., Papadopoulos, S., & Gennaro, C. (2024). MINTIME: multi-identity size-invariant video deepfake detection. *IEEE Transactions on Information Forensics and Security*.
- Concas, S., La Cava, S. M., Casula, R., Orru, G., Puglisi, G., & Marcialis, G. L. (2024). Quality-based artifact modeling for facial deepfake detection in videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3845-3854).
- Dang, H., Liu, F., Stehouwer, J., Liu, X., & Jain, A. K. (2020). On the detection of digital face manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5781-5790).
- De Lima, O., Franklin, S., Basu, S., Karwoski, B., & George, A. (2020). Deepfake detection using spatiotemporal convolutional networks. *arXiv preprint arXiv:2006.14749*.
- Do, N. T., Na, I. S., & Kim, S. H. (2018). Forensics face detection from GANs using convolutional neural network. *ISITC*, 2018, 376-379.
- Dörr, K. N., & Hollnbuchner, K. (2017). Ethical challenges of algorithmic journalism. *Digital journalism*, 5(4), 404-419.
- Dufour, N., Pathak, A., Samangouei, P., Hariri, N., Deshetti, S., Dudfield, A., ... & Bregler, C. (2024). Ammeba: A large-scale survey and dataset of media-based misinformation in-the-wild. *arXiv preprint arXiv:2405.11697*.
- Dwivedi, S., & Pachauri, S. DIGITAL REPOSITORIES: MANAGING, STORING AND DISSEMINATING DIGITAL CONTENT. *LIS TODAY*, 35.
- Farid, A. S. (2023). Changing the paradigm of traditional journalism to digital journalism: Impact on professionalism and journalism credibility. *Journal International Dakwah and Communication*, 3(1), 22-32.
- Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A., & Kashef, R. (2020). Recommendation systems: Algorithms, challenges, metrics, and business opportunities. *applied sciences*, 10(21), 7748.
- Fei, J., Xia, Z., Tondi, B., & Barni, M. (2022, December). Supervised gan watermarking for intellectual property protection. In *2022 IEEE International Workshop on Information Forensics and Security (WIFS)* (pp. 1-6). IEEE.
- Fernandes, S., Raj, S., Ortiz, E., Vintila, I., Salter, M., Urosevic, G., & Jha, S. (2019). Predicting heart rate variations of deepfake videos using neural ode. In *Proceedings of the IEEE/CVF international conference on computer vision workshops* (pp. 0-0).
- Fernando, T., Fookes, C., Denman, S., & Sridharan, S. (2019). Exploiting human social cognition for the detection of fake and fraudulent faces via memory networks. *arXiv preprint arXiv:1911.07844*.
- Frank, J., Eisenhofer, T., Schönherr, L., Fischer, A., Kolossa, D., & Holz, T. (2020, November). Leveraging frequency analysis for deep fake image recognition. In *International conference on machine learning* (pp. 3247-3258). PMLR.
- Gaillard, S., Oláh, Z. A., Venmans, S., & Burke, M. (2021). Countering the cognitive, linguistic, and psychological underpinnings behind susceptibility to fake news: A review of current literature with special focus on the role of age and digital literacy. *Frontiers in Communication*, 6, 661801.
- Galvan, F., Puglisi, G., Bruna, A. R., & Battiato, S. (2013). First quantization coefficient extraction from double compressed JPEG images. In *Image Analysis and Processing—ICIAP 2013: 17th International Conference, Naples, Italy, September 9-13, 2013. Proceedings, Part I* 17 (pp. 783-792). Springer Berlin Heidelberg.
- Gao, J., Xia, Z., Marcialis, G. L., Dang, C., Dai, J., & Feng, X. (2024a). DeepFake detection based on high-frequency enhancement network for highly compressed content. *Expert Systems with Applications*, 249, 123732.
- Gao, X., Wang, X., Chen, Z., Zhou, W., & Hoi, S. C. (2024b). Knowledge enhanced vision and language model for multi-modal fake news detection. *IEEE Transactions on Multimedia*.
- Ghai, A., Kumar, P., & Gupta, S. (2024). A deep-learning-based image forgery detection framework for controlling the spread of misinformation. *Information Technology & People*, 37(2), 966-997.
- Guarnera, L., Giudice, O., & Battiato, S. (2020). Deepfake detection by analyzing convolutional traces. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 666-667).
- Guo, Z., Yang, G., Chen, J., & Sun, X. (2020). Fake face detection via adaptive residuals extraction network. *arXiv preprint arXiv:2005.04945*.
- Hassan, N., Arslan, F., Li, C., & Tremayne, M. (2017, August). Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1803-1812).
- Hernandez-Ortega, J., Tolosana, R., Fierrez, J., & Morales, A. (2020). Deepfakeson-phys: Deepfakes detection based on heart rate estimation. *arXiv preprint arXiv:2010.00400*.
- Herzog, L. (2022). Shared standards versus competitive pressures in journalism. *Journal of Applied Philosophy*, 39(3), 393-406.
- Hu, J., Liao, X., Wang, W., & Qin, Z. (2021a). Detecting compressed deepfake videos in social networks using frame-temporality

- two-stream convolutional network. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3), 1089-1102.
- Hu, S., Li, Y., & Lyu, S. (2021b, June). Exposing GAN-generated faces using inconsistent corneal specular highlights. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2500-2504). IEEE.
- Huang, Y., Juefei-Xu, F., Guo, Q., Liu, Y., & Pu, G. (2022). Fakelocator: Robust localization of gan-based face manipulations. *IEEE Transactions on Information Forensics and Security*, 17, 2657-2672.
- Irshad, S., & Soomro, T. R. (2018). Identity theft and social media. *International Journal of Computer Science and Network Security*, 18(1), 43-55.
- Jastaniyah, A., & Bach, C. (2017). The importance of multimedia in information revolution. *Saudi Journal of Engineering and Technology*, 2(2), 89-99.
- Jeon, H., Bang, Y., & Woo, S. S. (2020, September). Fdftnet: Facing off fake images using fake detection fine-tuning network. In *IFIP international conference on ICT systems security and privacy protection* (pp. 416-430). Cham: Springer International Publishing.
- Kade, D. (2015). Ethics of virtual reality applications in computer game production. *Philosophies*, 1(1), 73-86.
- Karim, F., Oyewande, A. A., Abdalla, L. F., Ehsanullah, R. C., & Khan, S. (2020). Social media use and its connection to mental health: a systematic review. *Cureus*, 12(6).
- Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*.
- Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4401-4410).
- Khalid, H., & Woo, S. S. (2020). Oc-fakedect: Classifying deepfakes using one-class variational autoencoder. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 656-657).
- Kingma, D. P., & Welling, M. (2013, December). Auto-encoding variational bayes.
- Kitchens, B., Johnson, S. L., & Gray, P. (2020). Understanding echo chambers and filter bubbles: The impact of social media on diversification and partisan shifts in news consumption. *MIS quarterly*, 44(4).
- Koopman, M., Rodriguez, A. M., & Geradts, Z. (2018, August). Detection of deepfake video manipulation. In *The 20th Irish machine vision and image processing conference (IMVIP)* (pp. 133-136).
- Korshunov, P., & Marcel, S. (2018). Deepfakes: a new threat to face recognition? assessment and detection. *arXiv preprint arXiv:1812.08685*.
- Kumari, R., Gupta, V., Ashok, N., Ghosal, T., & Ekbal, A. (2024). Emotion aided multi-task framework for video embedded misinformation detection. *Multimedia Tools and Applications*, 83(12), 37161-37185.
- Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094-1096.
- Li, H., Li, B., Tan, S., & Huang, J. (2020). Identification of deep network generated images using disparities in color components. *Signal Processing*, 174, 107616.
- Li, Y., & Lyu, S. (2018). Exposing deepfake videos by detecting face warping artifacts. *arXiv preprint arXiv:1811.00656*.
- Li, Y., Chang, M.-C. and Lyu, S. (2018) In *Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking*. 2018 IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, 11-13 December 2018, 1-7.
- Liu, H., Li, X., Zhou, W., Chen, Y., He, Y., Xue, H., ... & Yu, N. (2021). Spatial-phase shallow learning: rethinking face forgery detection in frequency domain. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 772-781).
- Liu, Z., Qi, X., & Torr, P. H. (2020). Global texture enhancement for fake face detection in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8060-8069).
- López García, C., Sánchez Gómez, M. C., & García-Valcárcel Muñoz-Repiso, A. (2020, October). Scales for measuring Internet Addiction in Covid-19 times: Is the time variable still a key factor in measuring this addiction?. In *Eighth international conference on technological ecosystems for enhancing multiculturalism* (pp. 600-604).
- Lucas Graves, (2018). Understanding the promise and limits of automated fact-checking. Reuters Institute for the Study of Journalism.
- Lyu, S. (2024). DeepFake the menace: mitigating the negative impacts of AI-generated content. *Organizational Cybersecurity Journal: Practice, Process and People*.
- Maji, S., & Abhiram, A. H. (2023). "Mental health cost of internet": A mixed-method study of cyberbullying among Indian sexual minorities. *Telematics and Informatics Reports*, 10, 100064.
- Malik, A., Kuribayashi, M., Abdullahi, S. M., & Khan, A. N. (2022). DeepFake detection for human face images and videos: A survey. *IEEE Access*, 10, 18757-18775.
- Marra, F., Gragnaniello, D., Verdoliva, L., & Poggi, G. (2019, March). Do gans leave artificial fingerprints?. In *2019 IEEE conference on multimedia information processing and retrieval (MIPR)* (pp. 506-511). IEEE.
- Martyn, P. H. (2009). The Mojo in the third millennium: Is multimedia journalism affecting the news we see?. *Journalism Practice*, 3(2), 196-215.
- Masi, I., Killekar, A., Mascarenhas, R. M., Gurudatt, S. P., & AbdAlmageed, W. (2020). Two-branch recurrent network for isolating deepfakes in videos. In *Computer vision—ECCV 2020: 16th European conference, glasgow, UK, August 23–28, 2020, proceedings, part VII 16* (pp. 667-684). Springer International Publishing.
- Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023). Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Applied intelligence*, 53(4), 3974-4026.
- McCloskey, S., & Albright, M. (2019, September). Detecting GAN-generated imagery using saturation cues. In *2019 IEEE international conference on image processing (ICIP)* (pp. 4584-4588). IEEE.
- McGaughey, R. E. (2001). Application of Multimedia in Agile Manufacturing. In *Elsevier eBooks* (pp. 279-295).
- Meyers, A. (2000, December). Bobb Goldsteinn Coins the Term Multimedia: History of Information. Retrieved from History of Information Institute <https://www.historyofinformation.com/detail.php?id=2629>

- Michailidou, A., & Trenz, H. J. (2021). Rethinking journalism standards in the era of post-truth politics: from truth keepers to truth mediators. *Media, Culture & Society*, 43(7), 1340-1349.
- Millière, R. (2022). Deep learning and synthetic media. *Synthese*, 200(3), 231.
- Mittal, T., Bhattacharya, U., Chandra, R., Bera, A., & Manocha, D. (2020). Emotions don't lie: a deepfake detection method using audio-visual affective cues. In *Proceedings of the 28th ACM International Conference on Multimedia*. ACM (pp. 2823-2832).
- Nataraj, L., Mohammed, T. M., Chandrasekaran, S., Flenner, A., Bappy, J. H., Roy-Chowdhury, A. K., & Manjunath, B. S. (2019). Detecting GAN generated fake images using co-occurrence matrices. *arXiv preprint arXiv:1903.06836*.
- Nekmat, E., (2020). Nudge effect of fact-check alerts: Source influence and media skepticism on sharing of news misinformation in social media. *Social Media+ Society*, 6(1), 2056305119897322.
- Nguyen, H. H., Fang, F., Yamagishi, J., & Echizen, I. (2019, September). Multi-task learning for detecting and segmenting manipulated facial images and videos. In *2019 IEEE 10th international conference on biometrics theory, applications and systems (BTAS)* (pp. 1-8). IEEE.
- Nirkin, Y., Wolf, L., Keller, Y., & Hassner, T. (2020). Deepfake detection based on the discrepancy between the face and its context. *arXiv preprint arXiv:2008.12262*.
- Oyinloye, O., Akinola, R. A., Opaleke, E. A., & Okunade, S. J. (2024). Investigating the Impact of News Commercialization on Journalistic Ethics and Audience Trust. *African Journal of Social and Behavioural Sciences*, 14(2).
- Pal, A., Chua, A. Y., & Goh, D. H. L. (2020). How do users respond to online rumor rebuttals?. *Computers in Human Behavior*, 106, 106243.
- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- Pavlik, J. V. (2023). Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism & mass communication educator*, 78(1), 84-93.
- Pepple, I. I., & Acholonu, I. J. (2018). Media ethics as key to sound professionalism in Nigerian journalism practice. *Journalism and Mass Communication*, 8(2), 56-67.
- Qi, H., Guo, Q., Juefei-Xu, F., Xie, X., Ma, L., Feng, W., ... & Zhao, J. (2020, October). Deephythm: Exposing deepfakes with attentional visual heartbeat rhythms. In *Proceedings of the 28th ACM international conference on multimedia* (pp. 4318-4327).
- Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.
- Reviglio, U., & Agosti, C. (2020). Thinking outside the black-box: The case for "algorithmic sovereignty" in social media. *Social media+ society*, 6(2), 2056305120915613.
- Royakkers, L., Timmer, J., Kool, L., & Van Est, R. (2018). Societal and ethical issues of digitization. *Ethics and Information Technology*, 20, 127-142.
- Ruggiero, C., Karadimitriou, A., Lo, W. H., Núñez-Mussa, E., Bomba, M., & Sallusti, S. (2022). The professionalisation of journalism: Global trends and the challenges of training and job insecurity.
- Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., & Natarajan, P. (2019). Recurrent convolutional strategies for face manipulation detection in videos. *Interfaces (GUI)*, 3(1), 80-87.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001, April). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web* (pp. 285-295).
- Shabahang, R., Shim, H., Aruguete, M. S., & Zsila, Á. (2024). Oversharing on social media: Anxiety, attention-seeking, and social media addiction predict the breadth and depth of sharing. *Psychological reports*, 127(2), 513-530.
- SHAHBODIN, F., MOHD, C., RAHIM, N. R., PENDIT, U. C., DEWI, L., & KASIM, H. M. (2024). Systematic literature review on ethical consideration in multimedia professional practices. *Journal of Theoretical and Applied Information Technology*, 102(12).
- Shirky, C. (2013). Truth Without Scarcity, Ethics Without Force. In *The New Ethics of Journalism: Principles for the 21st Century* SAGE/CQ Press.
- Slater, M., Gonzalez-Liencre, C., Haggard, P., Vinkers, C., Gregory-Clarke, R., Jelley, S., ... & Silver, J. (2020). The ethics of realism in virtual and augmented reality. *Frontiers in Virtual Reality*, 1, 1.
- Songsri-in, K., & Zafeiriou, S. (2019). Complement face forensic detection and localization with facial landmarks. *arXiv preprint arXiv:1910.05455*.
- Tariq, S., Lee, S., & Woo, S. S. (2020). A convolutional lstm based residual network for deepfake video detection. *arXiv preprint arXiv:2009.07480*.
- Thomson, T. J., Angus, D., Dootson, P., Hurcombe, E., & Smith, A. (2022). Visual mis/disinformation in journalism and public communications: Current verification practices, challenges, and future opportunities. *Journalism Practice*, 16(5), 938-962.
- Wan, H., Luo, M., Ma, Z., Dai, G., & Zhao, X. (2024). How Do Social Bots Participate in Misinformation Spread? A Comprehensive Dataset and Analysis. *arXiv preprint arXiv:2408.09613*.
- Wang, G., Zhou, J., & Wu, Y. (2020a). Exposing deep-faked videos by anomalous co-motion pattern detection. *arXiv preprint arXiv:2008.04848*.
- Wang, R., Juefei-Xu, F., Ma, L., Xie, X., Huang, Y., Wang, J., & Liu, Y. (2019). Fakespotter: A simple yet robust baseline for spotting ai-synthesized fake faces. *arXiv preprint arXiv:1909.06122*.
- Wang, S. Y., Wang, O., Zhang, R., Owens, A., & Efros, A. A. (2020b). CNN-generated images are surprisingly easy to spot... for now. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8695-8704).
- Wu, X., Xie, Z., Gao, Y., & Xiao, Y. (2020, May). Sstnet: Detecting manipulated faces through spatial, steganalysis and temporal features. In *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 2952-2956). IEEE.
- Wu, Y., AbdAlmageed, W., & Natarajan, P. (2019). Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9543-9552).
- Xuan, X., Peng, B., Wang, W., & Dong, J. (2019, October). On the generalization of GAN image forensics. In *Chinese conference on biometric recognition* (pp. 134-141). Cham: Springer International Publishing.
- Yang, X., Li, Y., & Lyu, S. (2019a, May). Exposing deep fakes

- using inconsistent head poses. In ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 8261-8265). IEEE.
- Yang, X., Li, Y., Qi, H., & Lyu, S. (2019b, July). Exposing GAN-synthesized faces using landmark locations. In Proceedings of the ACM workshop on information hiding and multimedia security (pp. 113-118).
- Youvan, D. C. (2024). The Evolution of US Mainstream Media Headlines: From Investigative Journalism to Sensationalism in the Digital Age.
- Yu, N., Skripniuk, V., Abdelnabi, S., & Fritz, M. (2021). Artificial fingerprinting for generative models: Rooting deepfake attribution in training data. In Proceedings of the IEEE/CVF International conference on computer vision (pp. 14448-14457).
- Yu, Y., Ni, R., & Zhao, Y. (2020). Mining generalized features for detecting ai-manipulated fake faces. arXiv preprint arXiv:2010.14129.
- Zhang, X., Karaman, S., & Chang, S. F. (2019, December). Detecting and simulating artifacts in gan fake images. In 2019 IEEE international workshop on information forensics and security (WIFS) (pp. 1-6). IEEE.
- Zhang, Y., Zheng, L., & Thing, V. L. (2017, August). Automated face swapping and its detection. In 2017 IEEE 2nd international conference on signal and image processing (ICSIP) (pp. 15-19). IEEE.