Kabarak Journal of Research & Innovation Vol. 2025 No. Supplement 1 (2025)
ISSN 2305-784X (print) ISSN 2410-8383 (online)
https://journals.kabarak.ac.ke/index.php/kjri/

Special Issue Sponsored by

Artificial Intelligence and the battle for authenticity: safeguarding truth in the Digital Age

Author
Everlyn Chelangát Kimibei
Kisii University

Email: ekimibei@kisiiuniversity.ac.ke

Article History: *Submitted: 16th October 2025; Accepted: 9th February 2026; Published (online): 19th August 2026*

Abstract

Artificial intelligence (AI) has become a double-edged force in contemporary mass communication, enabling media innovation while accelerating the production and circulation of synthetic misinformation, particularly deepfakes. This paper examines the paradoxical role of AI in both facilitating and countering AI-generated disinformation, with a focus on its implications for public trust, media ethics, and democratic communication. Grounded in Media Literacy Theory, the study adopts a qualitative document analysis of academic literature, policy reports, technical studies, and verified media cases, including evidence from Kenya's 2022 general elections and comparable Sub-Saharan African contexts. The analysis shows that although AI-based detection and forensic tools offer important mitigation potential, their effectiveness is constrained by data bias, limited regional applicability, and unresolved ethical and regulatory challenges. In response, the paper advances a conceptual framework for AI-mediated authenticity, which conceptualizes authenticity as a negotiated outcome shaped by the interaction between AI generation technologies, detection systems, media literacy capacities, ethical norms, and policy environments. By foregrounding African democratic and media contexts, the study contributes an Africa-centered theoretical perspective to global debates on deepfakes and proposes practical recommendations for policymakers, educators, media practitioners, and technology developers aimed at safeguarding truth in the digital age.

Keywords: *Artificial Intelligence, Deepfakes, Misinformation, Media Ethics, Kenya, Media Literacy, Detection Tools*

1. INTRODUCTION

The influence of Artificial Intelligence (AI) is closely tied to the recent advancements in mass communication. Automated generation of content, tailored news feeds, AI-driven tools for public relations, etc. are some of the advancements in mass communication. Despite the positive impact of the mentioned technologies, they contribute to the increased distortion of reality in media. Deepfakes are examples of this. These AI media tools are capable of creating highly sophisticated audio-visual content that may be confused with real content (Citron & Chesney, 2019).

The rapid advancement of some AI tools causes an increase in disorder in the information. The disorder leads to the erosion of trust in journalism, creates the risk of media ethics violations, undermines the public discourse of democracy, etc. The impact of the same issues is also visible in Kenya and in the Sub-Saharan Africa region. There is the widespread use of manipulated videos and videos that are AI-generated in political contexts, such as on the platforms of Facebook and WhatsApp (Owino, 2022). There is a lack of awareness in detection and counteraction to the threats even in educated societies.

Concurrently, AI presents opportunities such as forensic watermarking, deepfake detection using neural networks, and real-time media authentication (Amerini et al., 2025). However, they remain underdeveloped due to technological, awareness, and content moderation/privacy challenges (Qureshi et al., 2024).

This paper explores the paradoxical role that AI play by being both a facilitator and counterforce of misinformation. Grounded in Media Literacy Theory, the paper aims to: (i) Examine how AI-generated content contributes to misinformation and public deception; (ii) Analyze the effectiveness and limitations of AI-driven detection and prevention tools; (iii) Propose ethical and educational strategies to foster media integrity and safeguard truth.

This paper engages with the conference theme on the balancing act of innovation and the ethics of the media by proposing the first of its kind contextually grounded, adaptable model for AI-mediated authenticity. In most studies, the authors and scholars define media authenticity as an inherent property of media content, whereas the current study defines it as a result of a negotiated process of the interplay of AI generation technologies, AI detection technologies, media literacy, ethics and regulation frameworks. With a focus on the African democratic and media contexts, the

model illustrates how the absence of data, lopsided digital skill sets, and regulatory deficits within Sub-Saharan Africa, places the region within the margin of the 'AI-optimized' misinformation economy. The current study modifies and contextualizes African Media Literacy Theory for an environment supersaturated with AI technologies and frameworks and offers an Africa-centric perspective on the dynamics of trust within and outside the system of contemporary digital media, and in how such trust is built, lost, and re/constructed.

Research Questions

Grounded in Media Literacy Theory, this paper addresses the following research questions:

i. In what ways does AI-generated content contribute to misinformation and public deception within contemporary mass communication, particularly in African democratic contexts?
ii. How effective are existing AI-driven detection and prevention tools in identifying and mitigating deepfakes, and what limitations constrain their application in Sub-Saharan African media environments?
iii. How can ethical governance and media literacy education be leveraged to strengthen media integrity and safeguard truth in AI-mediated communication systems?

2. LITERATURE REVIEW

2.1 Deepfakes and Generative Adversarial Networks (GANs)

The emergence of deepfakes is closely tied to the advancement of Generative Adversarial Networks (GANs); a class of machine learning models (Goodfellow et al., 2020). GAN consists of a generator and a discriminator that compete and refine until the generator can create realistic-looking synthetic material. Although GANs have valid applications in the medical field, entertainment, and virtual simulations, they have also been used to generate false media that questions the reliance on visual and audio data (Mirsky & Lee, 2020).

Deepfakes now have highly advanced and accessible tools. Non-experts can now create convincing manipulated videos using online tools and open-source frameworks like DeepFaceLab or FaceSwap, sparking widespread debates about authenticity in mass communication (Maras & Alexandrou, 2019). In recent times, however, these technologies contribute significantly to

information warfare and opinion control, such as during elections or during periods of war between countries.

The explanations provided in current scholarship on GANs tend to focus on the production of highly realistic synthetic media, yet much of this literature focuses on model performance and detection accuracy, and overlooks the social consequence . This has meant that deepfakes have been described almost exclusively as a technical problem, while overlooking the relevance of such technologies to (political) communication, trust, and media systems in the Global South. This focus potentially neglects the underrepresented fragility of African media systems, where disinformation generated by artificial intelligence could have greater social harm owing to little verification systems and patchy digital literacy.

2.2 Global and African Case Examples

Deepfakes have often been used around the world to mimic the voices and appearances of politicians, celebrities and company leaders. A fraudulent video of U.S. House Speaker Nancy Pelosi that misrepresented her as drunk was the first to raise concerns about the impact of video manipulation on political processes (West, 2017).

Disinformation aided by artificial intelligence begins to emerge in Africa around election time. During Kenya's 2022 general elections, presidential aspirants were targets of deep fake videos and altered images that circulated on WhatsApp and Facebook (Owino, 2022). According to the Mozilla Foundation's post-election evaluative reports, the materials, which were politically motivated, were often the result of collaborative efforts and were designed to achieve community polarization (Odanga, 2022).. In Nigeria and Ethiopia, similar manipulated content has been documented that exacerbates inter-ethnic conflicts, challenges to the legitimacy of the elections, and violence against the press (Wasserman & Madrid-Morales, 2022).

While international case studies highlight the disruptive possibilities of deepfakes at sensitive political moments, African-centric studies have been more sparse and of a more episodic nature. Most existing studies seem to recount instances of disconnected events and largely fail to provide sustained theoretical accounts of the role of misinformation in African elections and the various cultures and media of the continent. The absence of such studies in the African context speaks to the demand for more holistic approaches that go beyond descriptive case studies to analyze the

interrelation of the socio-technical and the institutional in creating the continental as well as the collective regional susceptibilities to AI-generated misinformation.

2.3 Theoretical Grounding: Media Literacy Theory

The theory of Media Literacy gives us an important way that help in understanding as well as responding to deep fakes. At its core, the theory promotes the ability to access, analyze, evaluate, and create media in various forms (Potter, 2010). It enables users to check the credibility of news, watch for deception and see the difference between truth and false information.

In today's world of AI, we need to teach people how to recognize algorithms and apply digital forensics. Scholars argue that education systems must now teach how AI algorithms curate content and how synthetic media can distort reality (Hobbs & Mihailidis, 2019). Since digital literacy in sub-Saharan Africa isn't universal, teaching people critical media skills is especially necessary to protect communities from misinformation.

Although Media Literacy Theory provides the basis for understanding how individuals can engage with the content of the media critically, most of the applications of the theory lack consideration of the generative AI technologies. The AI technologies are, for most intents and purposes, new. The older media literacy frameworks (which are still, ironically, current for much of the AI applications) emphasize interpreting messages, evaluating sources, and not so much understanding the role of the algorithm, AI, and synthetic media. The void calls for some refocusing of the media literacy frameworks, particularly in the political communication and digital infrastructure shallow-deepfake exposure interzone, the AI-mediated environments, and the deepfake exposure political communication zones.

2.4 Ethical Implications in an AI-Mediated Media Landscape

Such misinformation brings about further ethical problems, not just related to the detection process. Deepfakes introduce a new challenge to the right to privacy, consent, who has and who has not been involved and the psychology involved in the effects of deepfakes. The victims of non-consensual synthetic media, especially the women who fall victim of deep fake pornography, are in most cases damaged in reputation with little options of legal intervention (Citron & Chesney, 2019).

People cannot differentiate between truth and lies because of the synthetic content, so the activity of the press which is the defender of justice becomes more complicated. Should such tools be used improperly or deployed in a manner that discriminates the process, they may be violating the right of individuals to express their opinions. Consequently, media accountability and ethically driven innovation remain an enduring challenge that calls for multi-stakeholder partnerships across the state, technology, civil society, and the academy (UNESCO, 2021).

In addition, Global North legal and cultural presuppositions, which shape the ethical framing of deepfakes, lack cultural and contextual relevance in Africa's under-enforced regulatory, informal media economy, and digitally harmed victims' access to justice deficient frameworks.

3. METHODOLOGY AND APPROACH

This research utilizes qualitative conceptual analysis to examine the paradoxical role of artificial intelligence (AI) in facilitating and fighting misinformation, focusing primarily on deepfakes. The method combines academic literature, technical reports, media analysis, and case study materials to gain an overview of the contemporary digital landscape and its impact on veracity.

3.1 Research Design and Scope

The study adopts a multi-source document analysis method basing this on: Scholarly works on GANs, detection of deepfakes, and ethics of media (Maras & Alexandrou, 2019; Mirsky & Lee, 2020). Reports of policies and journalism from global and African contexts (UNESCO, 2021; Mozilla Foundation, 2023); and The publicly verifiable instances of media, including the 2022 elections of Kenya (Odanga, 2022; Owino, 2022).

The study of inclusion criteria necessitated that the documents offered the clear articulation of at least one of the following subjects: Artificial Intelligence (AI) generated synthetic media, deepfakes, methodologies of deepfake detection, ethics of the media, and misinformation in the digital world, and also the digital/media literacy. Also, documents had to be prioritized that 'specifically targeted' empirical evidence, a theoretical framework, or a policy component in the context of African or Global South media cultures. Opinions from editorial sources that lacked sufficient evidence and documents that had ambiguous authorship were also excluded.

In an effort to reduce selection bias, the study in the first place adopted a wide variety of perspectives, from the technological, educational, and ethical to journalistic and policy frameworks. At the same time, the study also incorporated African and non-African perspectives to address the concern. While primary data collection was not a primary objective of the study, the combination of contemporary and diverse documents to a great extent improved the transferability of the findings.

This approach enables the integration of an examination of both the technological dimensions (how disinformation generated by AI is created and how it is detected) and the social dimensions, particularly in developing democracies and changing media systems.

3.2 Theoretical Framework

This paper's primary theoretical framework involves Media Literacy Theory and the development of critical skills needed for the analysis of media and the subsequent evaluation and creation of media products (Potter, 2010). As for the context of AI, the skills further extend to the understanding of the algorithmic, the identification of the manipulative, and the evaluation of authenticity in media. It also responds to the need for the improvement of digital literacy, with the aim of equipping people with skills to combat disinformation (Hobbs & Mihailidis, 2019).

To enrich this framework, the study integrates AI ethics as a secondary lens, focusing on issues of consent, algorithmic accountability, and the ethical deployment of detection technologies (Citron, 2019; (Wasserman & Madrid-Morales, 2022).

3.3 Conceptual Framework for AI-Mediated Authenticity

Although the topic of deepfakes has been widely studied by scholars from a technological as well as an ethic point of view, the related theme of authenticity as a matter of AI-saturated communication forums, and specifically as an issue that concerns African democratic societies, has not received the attention it deserves. This research therefore bridges the gap by advancing a concept of AI-mediated authenticity.

The framework presents the notion of artificial intelligence as an entity that exists on two parallel streams. In the first stream, AI is positioned as the dispenser of disinformation as it relates to the generator of synthetic media that can be used to distort reality and compromise the credibility of

news sources in addition to manipulating public opinion. In the second stream, AI is positioned as a detector and mitigator.

However, the above-mentioned technological trends and challenges can be transformed and influenced in three important ways. First, the role of media literacy is important in determining the success or failure of misinformation generated using AI. Second, issues of ethics in the way that technologies of generative and detection systems can be considered legitimate and accepted. Third, the regulation of technologies that may cause harm can be limited or encouraged in environments where there is insufficient legal regulation of synthetic media.

Under this framework, authenticity is conceptualized not as an attribute of media content fixed by its producers, but as a result of social negotiation that arises from interaction between artificial intelligence networks, institutional protection, and civic empowerment. In shedding light on African electoral and media environments, this framework directs our recognition of how data disparities in sub-Saharan Africa, through both media regulation and digital competencies, heighten sensitivity to disinformation by artificial intelligence. This framework recasts Media Literacy Theory by contextualizing media interaction within an artificial intelligence-mediated information environment.

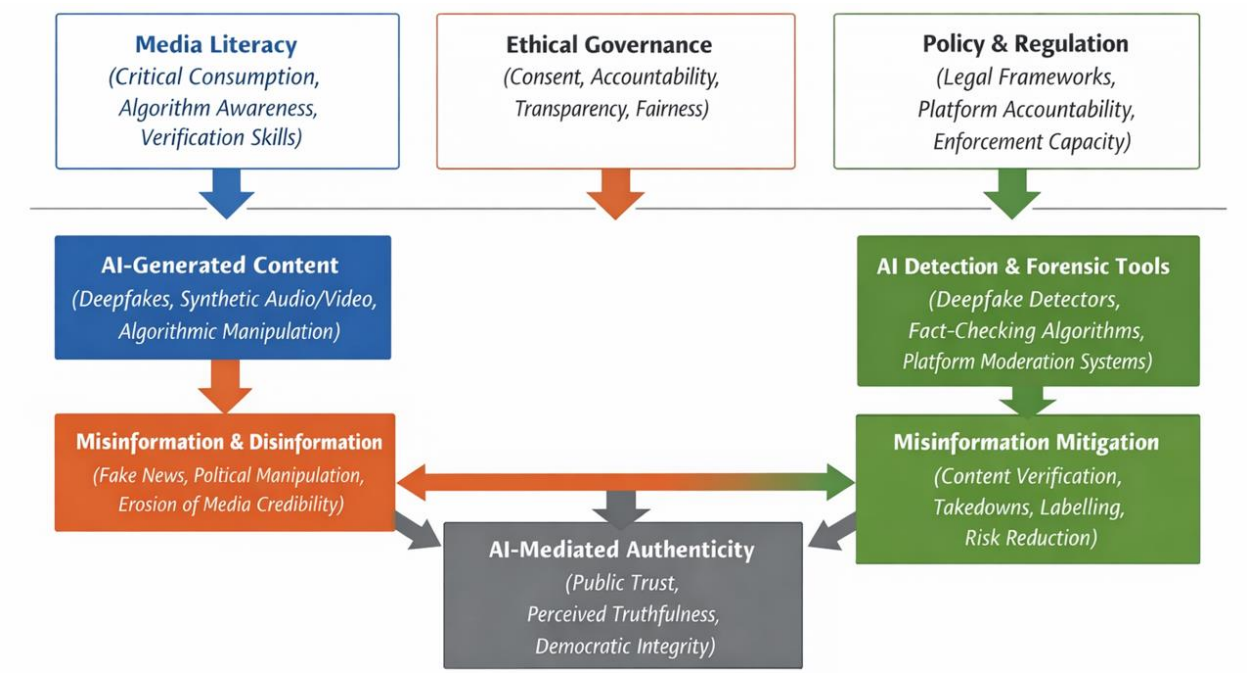

Figure 1: Conceptual Framework for AI-Mediated Authenticity

Conceptual model depicting AI as either the generator or detector of misinformation, moderated by media literacy, AI ethics, and AI policies affecting public trust and authenticity in digital communication.

3.4 Limitations

This study does not involve primary data collection. Due to the fact this is a conceptual paper, the sole reliance on secondary data entails a limitation on the understanding of localized, community-level impacts. That said, the chosen, recent, verifiable, and diverse representative materials do contribute to the credibility and contextual relevance of the findings. Furthermore, although primary focus is given to Kenya and the wider Sub-Saharan region, the majority of ethical and technical concerns are applicable across global media systems.

4. FINDINGS AND DISCUSSION

Employing the construct of authenticity in AI-mediated communication, as outlined in Figure 1, the findings will be structured around four interconnected themes focused on the processes of

digital authenticity through artificial intelligence: generation, mitigation, education, and regulation.

4.1 The Growing Threat of AI-Generated Disinformation

AI technologies, particularly Generative Adversarial Networks (GANs), are now advanced enough to create deepfakes that simulate real people's voices, faces, and mannerisms at a rapid rate. The political ramifications are substantial. One infamous global example is a deepfake of Ukrainian President Volodymyr Zelensky, which depicted him surrendering to Russia. Although this fabrication was promptly disproven, it was shared widely before it was removed (Allyn, 2022; Twomey et al., 2023).

AI-generated disinformation during the 2022 general elections in Kenya has localized consequences. Candidates were impersonated in deepfake videos, while other manipulated audio clips were attempts at ethnically divisive incitement (Owino, 2022). Due to the limited moderation of content in encrypted messaging apps such as WhatsApp, disinformation spread rapidly (Odanga, 2022). The disinformation campaigns were technologically advanced, coordinated, and sophisticated, and coupled with political actors and influencers.

The campaigns' erosion of public trust is widely documented. The Mozilla Foundation reports that 60 percent of Kenyan voters surveyed stated they had seen at least one misleading or modified video in the campaign (Odanga, 2022). These campaigns illustrate contexts in which AI is used to create fake events, as well as manufacture consent and other forms of social control in politically sensitive situations.

4.2 Tools and Techniques for Deepfake Detection

To combat the creation of synthetic media, AI-based detection systems have been implemented. Some of these systems consist of the following: the use of neural networks and the digital fingerprinting method from (Agarwal et al., 2020; Guarnera et al., 2020) which identifies visual inconsistencies (blinking patterns, lighting mistakes, etc.), audio mismatches, and digital fingerprinting. Google, Microsoft, and Facebook have each funded projects in the detection algorithm area and data sets such as FaceForensics++ and the Deepfake Detection Challenge (Kwan et al., 2024).

These tools also have a number of drawbacks. The detection methods and the generation methods have different paces of advancement and therefore act in a cat and mouse manner. As the world becomes more technologically advanced so do the deep fakes and therefore the harder they are to detect (Mirsky & Lee, 2020). The majority of detection models use western data and so the models are not as effective in the African media environments because they overlook the structure of African faces, the languages, and the cultures (Wasserman & Madrid-Morales, 2022).

Detection tools also are not used until the damage has already been done. Once a piece of content goes viral, detection tools flag the content and the damage is done. This damage highlights the need for more educational tools and ethical guidelines.

4.3 Education and Media Literacy as Long-Term Solutions

The problem of AI-generated disinformation cannot be approached using just tech-based solutions. In vulnerable democracies, the most effective long-term solutions will likely involve the teaching of media literacy. Media Literacy Theory proposes that people should be taught to go beyond the simple consumption of media, and that they should be taught how to question the sources, purposes, and legitimacy of the media (Potter, 2010).

In Kenya and Nigeria, integration of digital literacy into secondary school and journalism school curricula has begun (Hobbs & Mihailidis, 2019). Public information campaigns aimed at correcting misinformation and fact-checking toolkits have been created by the NGOs Code for Africa and Africa Check.

These initiatives have been implemented in a patchy and underfunded manner. The systemic approach needed in this case would be one of policy, in which the embedding of media literacy in school curricula becomes a requirement, tech companies are made to increase transparency, and funding directed to public-interest journalism to counter the narratives of disinformation (UNESCO, 2021).

4.4 Ethical and Policy Gaps

Issues brought about by the advent of AI-generated misinformation go beyond mere technological concerns. People are expressing worries about privacy, consent, and the subsequent impact on the right to free speech. For example, deepfake pornography is an imbalanced phenomenon that primarily targets women and other marginalized communities, and in many places, there are even

no laws to provide oversight due to the absence of specific cybercrime legislation (Citron & Chesney, 2019).

Laws on synthetic media are, to varying degrees of completion, only present in a handful of African countries. The Computer Misuse and Cybercrimes Act of 2018 in Kenya has provisions that broadly cover the phenomenon of 'fake news' but does not cover news that is AI-generated. This in-between state of the law is a hurdle to prosecution and enforcement, thereby sustaining the impunity of the 'creators' and 'distributors' of damaging synthetic media (Wasserman & Madrid-Morales, 2022).

That is why policymakers should change their frameworks to address the special risks from AI-based misinformation. It means outlining deepfakes in legal terms, holding people responsible for their misuse and supporting morally responsible progress.

5. CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

Artificial intelligence is now quickly changing how information is generated, shared and viewed in mass communication. Generative models such as GANs can make processes more efficient and drive new ideas, but they have also helped bring about a new wave of disinformation. As a large number of people now turn to deepfakes and other artificial content, fake news poses a threat to realism, social trust, and political health in those regions still in the process of becoming digitally literate.

This paper has provided a survey of how AI can serve as a deterrent and solution to, misinformation. With reference to Kenya and Sub-Saharan Africa, as Media Literacy Theory and international examples demonstrate, deepfakes have become a tangible threat. They assist in manipulating individuals in elections, the conflict of ethnic groups and social campaigns.

These tools enhance the possibility to address the problem of synthetic content. However, the weaknesses of such systems such as a regional bias in the data, outdated technology and ineffective policing show that technology cannot fix the issue on its own. The problem of privacy, consent of users and content management make the implementation of these tools more complicated. That is

why we need the strategies that combine the technology development, ethical rules, and the education of the people.

5.2. Recommendations

Since the major theme of the conference was the need to balance innovation and ethics in media and public relations, the following recommendations were proposed:

To Policymakers: Criminalize and clarify that it is unlawful to use deepfakes to do harmful actions, such as impersonate politicians or release fake images without their consent; Open and make social media and AI creators accountable; and Organize a task force of media, civil society, academia and tech companies to manage and solve disinformation issues.

For Educators and Schools: Make sure media and algorithmic literacy is part of the national curriculum by teaching students how to spot, study and doubt any digital content; Join forces with NGOs and organizations that fact check to hold training sessions for people in the community to detect false news.

For Media Practitioners and Journalists: Review user-provided and AI-generated content to make sure it is real before you post or share it; Technology teams should be included in planning to automatically spot fake news in newsrooms; and Help educate the public about deepfakes and what problems synthetic media may cause.

For AI Developers and Technology Companies: AI detection tools should be educated using data from across different regions to suit African situations; Ethical ideas such as explainability, fairness and responsibility, must be part of designing algorithms; Help support grassroots learning on digital security by making current detection tools available to anyone under an open-source license.

Policy improvements, more education, leading by example and ethical progress can all increase community resistance to synthetic disinformation. The main challenge in digital authenticity today is about morals and maintaining truth, trust and responsible government.

Funding Statement

This research received no external funding.

Conflict of Interest Statement

The authors declare that there is no conflict of intertest.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

Agarwal, S., Farid, H., Fried, O., & Agrawala, M. (2020). Detecting deep-fake videos from phoneme-viseme mismatches. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2020-June*, 2814–2822. https://doi.org/10.1109/CVPRW50498.2020.00338

Allyn, B. (2022). *A deepfake video showing Volodymyr Zelenskyy surrendering worries experts : NPR*. https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia

Amerini, I., Barni, M., Battiato, S., Bestagini, P., Boato, G., Bruni, V., Caldelli, R., De Natale, F., De Nicola, R., Guarnera, L., Mandelli, S., Majid, T., Marcialis, G. L., Micheletto, M., Montibeller, A., Orrù, G., Ortis, A., Perazzo, P., Puglisi, G., … Vitulano, D. (2025). Deepfake Media Forensics: Status and Future Challenges. *Journal of Imaging 2025, Vol. 11, Page 73*, *11*(3), 73. https://doi.org/10.3390/JIMAGING11030073

Citron, D. K., & Chesney, R. (2019). Deepfakes and the New Disinformation War. *Foreign Affairs*. https://scholarship.law.bu.edu/shorter_works/76

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144. https://doi.org/10.1145/3422622;SUBPAGE:STRING:BASIC

Guarnera, L., Giudice, O., & Battiato, S. (2020). DeepFake Detection by Analyzing Convolutional Traces. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2020-June*, 2841–2850. https://doi.org/10.1109/CVPRW50498.2020.00341

Hobbs, Renee., & Mihailidis, Paul. (2019). *The international encyclopedia of media literacy*. https://books.google.com/books/about/The_International_Encyclopedia_of_Media.html?id=Bt9NuwEACAAJ

Kwan, C., Li, B., Yu Gong, L., & Jun Li, X. (2024). A Contemporary Survey on Deepfake Detection: Datasets, Algorithms, and Challenges. *Electronics 2024, Vol. 13, Page 585*, *13*(3), 585. https://doi.org/10.3390/ELECTRONICS13030585

Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *International Journal of Evidence and Proof*, *23*(3), 255–262. https://doi.org/10.1177/1365712718807226;ISSUE:ISSUE:DOI

Mirsky, Y., & Lee, W. (2020). The Creation and Detection of Deepfakes: A Survey. *ACM Computing Surveys*, *54*(1). https://doi.org/10.1145/3425780

Odanga, M. (2022). *Opaque and Overstretched, Part II - Mozilla Foundation*. https://www.mozillafoundation.org/en/campaigns/opaque-and-overstretched-part-ii/

Owino, V. (2022). *Kenya election: Deep fakes, propaganda, libel inundate social media - The EastAfrican*. https://www.theeastafrican.co.ke/tea/news/east-africa/kenya-election-deep-fakes-propaganda-inundate-social-media-3904934

Potter, W. J. (2010). The State of Media Literacy. *Journal of Broadcasting & Electronic Media*, *54*(4), 675–696. https://doi.org/10.1080/08838151.2011.521462

Qureshi, S. M., Saeed, A., Almotiri, S. H., Ahmad, F., & Ghamdi, M. A. A. (2024). Deepfake forensics: a survey of digital forensic methods for multimodal deepfake identification on social media. *PeerJ Computer Science*, *10*, e2037. https://doi.org/10.7717/PEERJ-CS.2037

Twomey, J., Ching, D., Aylett, M. P., Quayle, M., Linehan, C., & Murphy, G. (2023). Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. *PLOS ONE*, *18*(10), e0291668. https://doi.org/10.1371/JOURNAL.PONE.0291668

UNESCO. (2021). *Threats that silence: Trends in the safety of journalists | World Trends in Freedom of Expression and Media Development: 2021/2022 Online Report*. https://www.unesco.org/reports/world-media-trends/2021/en/safety-journalists

Wasserman, Herman., & Madrid-Morales, Dani. (2022). *Disinformation in the global South*.

West, D. M. (2017). *How to combat fake news and disinformation | Brookings*. https://www.brookings.edu/articles/how-to-combat-fake-news-and-disinformation/