The Dual Impact of Artificial Intelligence on Media Credibility: Challenges, Opportunities and Strategies

Shengge Xu
College of Liberal Arts and Social Sciences, City University of Hong Kong, Hong Kong, China
shenggexu2-c@my.cityu.edu.hk

Abstract. This paper examines the impact of AI on media credibility through a review of the literature and qualitative case analyses of some current applications including NLP technology for disinformation tracking and deepfake detection.While AI is both a facilitator and obstacle for credible journalism, it significantly increases journalistic credibility through improved accuracy of content production, scalable fact checking, and providing tailored user experiences for media sustainability. While AI has many benefits, numerous systematic risks exist. Algorithmic biases reinforce social inequalities and amplify filter bubbles, leading to fragmentation of the public sphere and difficulties for intercommunity communication across the diversity of our society. Deep fakes that are indistinguishable from reality put the basis of evidence of journalism at risk and cause epistemic uncertainty. In addition, black box algorithms' opacity endangers transparency of media and the vital relationship of trust between media and audience. This paper argues that reliable AI governance grounded in journalistic professional ethics and democratic accountability is necessary. As such, this paper proposes a strategic multi-stakeholder public or private framework including 3 pillars: Implement human-in-the-loop auditing processes to ensure continuous editorial control; Establish an ethical AI guideline for the industry including transparency and content labeling demands; Promote public media literacy with focus on algorithmic awareness. All of the above will maximize the augmentative potential of AI while protecting media trust in an increasingly synthetic and algorithmically mediated information environment.

Keywords: Artificial Intelligence, Media Credibility, Algorithmic Bias, Deepfakes, Media Ethics.

1 Introduction

The integration of AI is radically transforming news production, distribution and consumption across all areas of the media landscape. Emerging technologies such as automated journalism and personalized recommendation systems provide unparalleled efficiencies, yet also introduce several significant risks including algorithmic biases, deepfake proliferation, and therefore directly threaten media credibility as a foundation for democratic deliberation. The ambiguous impact of AI on society requires systematic

G. E. Atalay and A. Wang (eds.), *Proceedings of the 2026 3rd International Conference on Public Relations and Media Communication (PRMC 2026)*, Advances in Social Science, Education and Humanities Research 1032,
https://doi.org/10.2991/978-2-38476-609-3_22

investigation to understand the multifaceted role of AI as a force for eroding and fostering public trust.

This study carries big meaning to technology ethics and computational journalism fields, for media workers and policy makers who take charge of responsible AI application and journalism honesty. Hence, this study will try to inspect artificial intelligence's influence on news media's credibility through three paths: chances for precise information, audience attraction, and fact-checking support; related dangers of biased information, synthetic media, and non-transparency; as well as ethics-related guidance. To grasp the complex connections between AI systems and media trust, the research will use an evidence-based method via wide-ranging literature review to combine existing knowledge. Therefore, the large objective of this methodology is to understand AI's two-sided character in media environments by looking into AI systems' enabling and destabilizing effects beyond a deterministic angle. Ultimately, this study will provide a multi-stakeholder viewpoint about AI's use as an augmentation technology for credible journalism, thus also making preparation for AI's risks to media trust.

2 Positive Impacts of AI on Media Credibility

AI Technology has great potential to improve the quality, efficiency and accuracy of media content while providing technical support for reconstructing and consolidating the credibility of media.

2.1 Enhancing Content Production Efficiency and Accuracy

AI-driven automated tools are fundamentally changing the way journalists work, which has a direct positive impact on the quality and accuracy of journalism. In structured forms of journalism such as financial reports, sports results, and election data, automated content generation can generate factual, clear narratives extremely quickly. Algorithmic journalism allows for quick reporting in breaking news situations and eliminates human errors when processing vast amounts of routine data [1]. AI can perform repetitive tasks, giving real journalists time for analysis, investigative work and nuanced storytelling; areas where human judgement and human empathy remain irreplaceable.

AI analytics are greatly improving investigative journalism. Machine learning algorithms can look at huge datasets such as public procurement records, corporate filings, satellite images and find patterns, anomalies, and links between things that would be impossible for humans to find by searching manually [2]. This has produced valuable understandings regarding corruption, the environment, and society. NLP makes possible the inquiry of thousands of documents to gain cognition about the spread of disinformation or the sponsors of influence campaigns . Therefore, AI provides factual foundation for journalism, thus the media can strengthen their argument to gain credibility[3]. Hence, improved AI-supported transcription and translation services for reporting on multiple geographic and linguistic communities will raise the possibilities for credible journalism.

2.2 Strengthening Personalized Recommendation and User Engagement

In a crowded digital economic environment, high-quality journalistic works have need of readers. AI recommendation systems which base on public interest can assist citizens to locate trustworthy and related information. AI algorithms that receive training from browsing behavior, reading behavior, and usage modes can provide a personalized news stream containing less well-known news and updates on topics which the user favors[4]. This kind of considerate personalization will restrict information overload and make users stay with high-quality journalistic works for more time, thus forming a kind of attachment feeling to high-quality journalism sources.

From a business viewpoint, sustainable reader loyalty is essential for the economic survival of journalism field, because content that matches readers' interest can bring about subscriptions and loyalty, not the unsteady advertising income or sensational news headlines[5]. In addition, only a good financial state can enable news rooms to produce high-quality journalistic works and maintain editorial independence, both of which are necessary components for long-term credibility. Therefore, ethical-based personalization will enhance reader experience and also promote the economic sustainability of credible journalistic works.

2.3 Assisting in Fact-Checking and Disinformation Identification

The proliferation of misinformation on the Internet is a risk for journalism's credibility and the quality of public debate. AI can help human fact checkers and journalists in this battle with scalable and powerful solutions.More advanced natural language processing models can be trained to identify the linguistic characteristics of false or misleading content-such as hyperbole, emotional appeals, contradiction, and commonly used propaganda techniques [3]. These systems can examine social media and news websites for potential misinformation in real time and identify potentially suspicious claims for human review.

AI is critical in verifying multimedia content. Image and video algorithms allow for reverse searching for the original document, assessing the metadata for signs of tampering, and detecting deep fake generation artefacts [6]. The use of network analysis algorithms can help identify the propagation routes for viral misinformation and identify bot networks and coordinated inauthentic behavior. AI can help verification desks and independent fact-checkers at news outlets by providing them with a force multiplier to fight against and debunk falsehoods more quickly and efficiently than ever before. In addition to correcting individual inaccuracies, the media will also be able to demonstrate a commitment to accuracy and will be reaffirmed as authoritative sources of truth through proactive defense of against falsehoods.

3 Challenges and Risks Posed by AI to Media Credibility

The same technologies which create solutions can be used to create deep, systemic risks that could erode public trust if left unmitigated.

3.1 Algorithmic Bias and the Information Cocoon Effect

AI systems are developed based on historical input data and thus reflect the biases that exist in society today. Therefore, when biased algorithms are used to create curation of news content, content moderation and predictive reporting, the algorithm provides favouritism to certain perspectives, sources and narratives while ignoring other ones [7].To illustrate, with engagement data, an algorithm might promote sensational and divisive content, since such posts get more clicks and rank untrustworthy sources higher than trustworthy news sources. Consequently, a "filter bubble" emerges where individuals are confronted solely with information that supports their preexisting beliefs [8]. The fragmented common information space does not fulfill the media's democratic ideal of providing a shared basis for public debate; it hardens polarization and makes an audience more prone to confirmation bias. The credibility of any source showing a different viewpoint dwindles. The public does not see media as a reflection of reality, but rather as an editor of algorithmically selected and filtered fragments of reality. Journalism suffers.

3.2 Deepfakes and the Proliferation of Synthetic Content

Generative AI models can create very realistic synthetic media ("deepfakes"). This poses an existential threat to the fabric of evidence in journalism.The ability to create a convincing video of a politician saying something they actually never said, or creating imagery for an event that never happened, creates a "liar's dividend" [9]. Malicious actors can use the existence of deepfake technology to cast doubts on real footage of events by claiming it is fake. This compromises the public's ability to trust visual or audio evidence — a basic tenet of today's news service.

There are two types of damage caused by a deepfake: immediate and long-term. Deepfakes can manipulate public opinion, disrupt elections, and incite violence quickly, but the growing prevalence of deepfakes is contributing to a long-term environment characterized by pervasive epistemic uncertainty or a state of nihilistic disbelief which is damaging to both media credibility and democracy. The number of resources necessary to debunk advanced synthetic media is great but news organizations have limited resources to address this challenge.

3.3 Lack of Transparency and the Crisis of Public Trust

Most complex AI systems can be called "black boxes".Their decisions about why one story got recommended and another did not, or why something was flagged as inappropriate usually cannot be understood by their engineers [10]. Journalism's ethics around accountability and transparency conflict directly with algorithms' opacity conflict.

When audiences lack transparency their news feed or moderation decisions are made by an online platform, it creates mistrust, cynicism, and possible allegations of secret or biased purposes. This loss of confidence hurts the relationship between media organisations and the institutional credibility between the media and the public. This situation becomes worse as the media industry does not clarify how they are using artificial intelligence in the production of news including writing an article, creating an image, or selecting a headline [5]. The absence of attribution blurs the line between

human and machine authorship and erodes trust because audiences feel misled about where their news comes from and how it is produced.

4 Strategic Recommendations for AI Application in Media

To address the impact of AI, the research will implement a multi-stakeholder approach to promoting ethical use, transparency, and empowerment.

4.1 Constructing a "Human-in-the-Loop" Content Auditing Mechanism

Developing a "Human-in-the-Loop" (HITL) audit process for all content AI can replace humans as the support organisation, using HITL to replace humans in editor and moderation decisions [11]. AI will be able to scan for misinformation as well as finding large quantities of data to identify trends and filter non-legal content through existing filters. All final editorial decisions, especially those that contain nuances, contexts or ethics and consideration of effects will be made by a qualified journalist or editor.

Collaborating to achieve news values, ethics, and cultural sensitivity allows for better decisions on significant issues made by news organisations. News organizations should have standards on AI/human collaboration in terms of procedures, human review criteria, and appeal procedures of algorithmic decisions. Standards should allow journalists to maintain editorial control but also allow for human oversight of public speech.

4.2 Establishing AI Ethical Guidelines and Transparency Standards

News media must take active action to compile their own AI ethics charter by themselves. The standards need to be open and show firm promise to core values, thus including auditing and de-biasing training data and outputs; human beings' accountability for AI-produced results; and educating the audience on how to carry out interpretation of AI application.

Transparency must be converted into concrete actions. Choices for algorithmic explainability should be provided to users, letting them know the reason for receiving certain recommendations, such as "This story is trending in your region" or "You are seeing this because you read about topic Y". At the very least, clear and eye-catching labeling is a necessary requirement. Any content generated or greatly changed by artificial intelligence should be labeled in this way [12]. Therefore, industry consortia ought to cooperate with academic researchers and civil society groups; hence, they can align disclosure practices and develop shared audit tools to assess algorithms' ethical performance.

4.3 Enhancing User Media Literacy and Fostering Participatory Oversight

Improving media credibility in AI epoch needs a citizen group with empowerment and critical literacy. Therefore, media enterprises will obtain benefits from next-generation media literacy programs that exceed simple source assessment. Thus, digital literacy can instruct citizens about algorithm operations like recommendation procedures and ranking systems, their hidden commercial and ideological driving forces, hence how to adjust personal settings to acquire alternative information resources [13].

Education might center its effort on instructing people to identify falsified media through lighting, shadows, audio-video synchronization and reverse image searching. Therefore, news institutions may construct trust relations with their audience by making audience participation in their reporting an institutionalized system. This system might contain channels for audience to put forward feedback on the making logic of their recommendation contents, a public editor who inspects the operation effect of algorithms, and crowd-participated projects to carry out analysis on the algorithms hence. Through working cooperatively, media organizations can strengthen transparency collaborative, mutually beneficial, and ultimately, more trustworthy relationships with their respective communities.

5 Conclusion

This research work probe into the positive and negative influence that AI brings for the improvement and trustworthiness enhancement of media. The research carry out analysis on four aspects: how AI can help media to become more trustworthy, how AI can give assistance to the engagement model, how to make media more in line with ethics, and how to fight against disinformation. Therefore, all these new technical applications will let media have more effectiveness and stronger influence. Regretfully, systemic risks still exist for all these new technical applications. Algorithmic bias could make social inequality continue to exist and divide the information space through filter bubbles, thus reducing the public democratic discussion. Hyper-realistic synthetic media brings fundamental threat to the evidence basis of journalism and cultivates public distrust toward information. Hence, many AI systems are black boxes, which violate the basic journalistic principles of transparency and accountability, and break the mutual agreement with the audience. If such systemic risks are not controlled, they will damage the credibility that AI tries to improve.

The outcome of this research is that ethical governance must be implemented for responsible AI-media behaviors, and such governance should take human beings as the core. The research put-forward multi-dimensional strategy plan includes three parts: development of "human-in-the-loop" auditing procedures, formulation of industry charters related to AI ethics and transparency that cover labeling work, and promotion of media literacy among the general public with focus on algorithm-related knowledge; hence, this plan can push related work forward. Therefore, AI can be placed in a position that does not replace but improves journalists' judgment ability, and apply technology to maintain professional ideals for the service of the public.

This research will provide actual beneficial values to academic circle, industrial field, and policy-making personnel. An analytical framework for future research about sociotechnical systems will be provided. Therefore, news organizations and technology companies must blend fairness, accountability, and explainability into design and deployment of their AI systems, as this study's results have evidenced. Meanwhile, policy makers and educators must make incentive measures for cultivating public's critical literacy and establish protective safeguards for the whole information ecosystem. Hence, future research work should explore long-term influence of AI on

trust, construct more expansive and transparent algorithm auditing software, and design innovative participatory mechanisms to let public join in governing media-assisting technologies. Thus, journalism's credibility will depend on continuous cooperation among stakeholders to guarantee technology's progress conforms with unchanging ethical values and serves society's democratic desires in an AI era.

References

1. Diakopoulos, N.: Automating the news: How algorithms are rewriting the media (2nd ed.). Harvard University Press (2023).
2. Broussard, M., Diakopoulos, N., Matthews, J.: Artificial intelligence and investigative reporting: New tools, new ethics. Journalism Practice 18(3), 345–362 (2024).
3. Shu, K., Mahudeswaran, D., Liu, H.: FakeNewsNet: A data repository with multi-modal detection of fake news. Proceedings of the International AAAI Conference on Web and Social Media 17, 1000–1008 (2023).
4. Kormelink, T.G.: Negotiating news relevance: How algorithmic and editorial curation shape news encounters. New Media & Society 25(12), 3354–3373 (2023).
5. Brennen, J.S., Howard, P.N., Nielsen, R.K.: What to expect when you're expecting robots: The professional logic of AI adoption in newsrooms. Digital Journalism 12(1), 1–20 (2024).
6. Westerlund, M.: The emergence of deepfake technology: A review. Technology Innovation Management Review 14(2), 4–16 (2024).
7. Noble, S.U.: Algorithms of oppression: How search engines reinforce racism (2nd ed.). New York University Press (2022).
8. Pariser, E.: Beyond the filter bubble: Concepts, myths, and evidence. The Annals of the American Academy of Political and Social Science 700(1), 45–58 (2024).
9. Chesney, R., Citron, D.: Deep fakes and the new disinformation war: The coming age of post-truth geopolitics. Foreign Affairs 102(1), 147–160 (2023).
10. Ananny, M., Crawford, K.: Seeing without knowing: The limits of algorithmic transparency. International Journal of Communication 16, 1225–1244 (2022).
11. Gillespie, T.: Do not recommend? Reduction as a form of content moderation. Social Media + Society 10(1) (2024).
12. Diakopoulos, N., Koliska, M.: Algorithmic transparency in the news media. Digital Journalism 11(5), 945–964 (2023).
13. Powers, E., Jones, M.: Teaching the machine: Integrating algorithmic literacy into media education curricula. Journal of Media Literacy Education 17(1), 22–35 (2025).