AI and Ethics: Moral Considerations of Automated Rumor Detection Technology

Yifei Deng[1,a,*]

[1]*Faculty of Humanities and Social Sciences: Globalization and Development, Beijing Normal University-Hong Kong Baptist University United International College, Guangdong, 519000, China*

a. t330032005@mail.uic.edu.cn

*corresponding author

Abstract: As digital society continues to develop, social media and information distribution have experienced significant changes, leading to challenges with accessing credible information and identifying misinformation. In response, the use of AI automation technology has been advocated as a solution to identifying fake news and improving information credibility. However, implementing this technology raises ethical concerns related to data privacy protection, transparency in information monitoring, and maintaining a balance between free speech and censorship. This paper examines ethics associated with using AI automation techniques for identifying misinformation, and proposing relevant laws and ethical guidelines to address these concerns. Through a literature review and case analysis, this study provides guidance for the ethical implementation of AI automation techniques for identifying misinformation. Our research aims to contribute to the development of sound ethical practices in the use of AI automation techniques in misinformation identification. Furthermore, this study highlights the need for an integrated approach in AI automation technology implementation and the significant role of transparency and accountability in mitigating ethical concerns. Additionally, this paper explores the potential benefits and pitfalls associated with using AI automation techniques to identify misinformation. While such technology can improve information credibility and protect the democratic process, it also poses risks to the quality of information and public trust in technology. In conclusion, this study underscores the importance of ethical considerations in implementing AI automation techniques for identifying misinformation and provides guidance for policymakers and practitioners working in the field of AI and information governance.

Keywords: digital society, social media, information dissemination, rumors, ethical considerations.

1. Introduction

With the rapid development of social media and information dissemination, the proliferation of rumors and false information has become increasingly rampant [1-3]. AI automation technology has been extensively utilized for rumor identification and information curation, providing novel insights to combat rumors and uphold information security [4]. However, the application of AI automation technology confronts ethical and legal issues, including concerns regarding privacy protection,

transparency in information monitoring, and the balance between free speech and censorship [4]. In recent years, the pervasiveness of misinformation has become a significant issue. Misinformation can arbitrarily sway people's opinions and can have devastating consequences, posing severe threats to social stability and national security. AI automation technology has become a powerful instrument to address the issue of misinformation [3]. Nevertheless, the adoption of AI automation technology raises many ethical and legal issues that cannot be ignored [4]. To ensure the ethical and legal implementation of AI automation technology for identifying rumors, it is essential to strengthen privacy protection while simultaneously developing transparent mechanisms for information monitoring [5]. Moreover, it is critical to maintain a balance between free speech and censorship, necessitating the establishment of distinct protocols to guide the application of AI automation technology [5]. In a nutshell, AI automation technology has illuminated innovative ideas for combating rumors and maintaining information security [3]. Nevertheless, it is crucial to take heed of ethical and legal considerations regarding its application. Appropriate laws and ethical guidelines must be established to ensure the ethical and lawful deployment of AI automation techniques for identifying rumors [4]. The objective of this article is to investigate these issues and propose pertinent legal regulations and ethical suggestions to serve as guidance for the ethical implementation of AI automation techniques for rumor identification.

2.   The Problem of Privacy Protection

The issue of privacy protection in AI has drawn widespread attention and concern [4]. With the development of a digital society, data collection and sharing have become an inevitable trend in social media and information dissemination. However, there are risks and challenges associated with the balance between data sharing and privacy protection [4]. Therefore, it is essential to optimize technical design and develop corresponding laws and regulations for the application of AI automation [4]. Moreover, data collection may leak personal privacy information [5], making it crucial to explore effective methods for data collection and privacy protection, particularly when gathering personal information [4]. In addition, data sharing can maximize the value of data but also pose an increased risk of privacy information leakage [4]. Authorization and protection issues need to be carefully considered in data-sharing practices, and relevant laws and regulations should be established [6]. The privacy protection concerns related to data collection and sharing necessitate the development of corresponding legal and ethical standards to ensure that the application of AI automation technology is more standardized and reliable [6].

Transparency in information censorship is an important topic that the public is concerned with AI research [7]. The transparency and fairness of the process are essential in rumor detection and information censorship [8]. The transparency of the information censorship mechanism is a key factor in ensuring its fairness and credibility. Therefore, during the information censorship process, AI automation technology needs to be transparent in its algorithms and decision-making processes to enhance its fairness and credibility [9]. The fairness and bias in information censorship should also be a concern as problems of individual and group biases exist. Therefore, it is necessary to enhance information diversity and establish a correction mechanism for biases in the design and practice of information censorship to increase its fairness and rigor [8]. The interpretability of algorithm decision-making processes should also be emphasized. To enhance the credibility and persuasiveness of the AI automation technology's algorithm decision-making, it needs to be interpretable to provide an explanation for why information is identified as true or false [9].

3.   The Way to Balance Civilized Censorship and Freedom of Speech

In the fight against rumors and information security, balancing civilized censorship and freedom of speech becomes an urgent problem that needs to be addressed. Information censorship must have social acceptance to guarantee its legality and morality [7]. Therefore, when designing and implementing information censorship, human factors need to be considered, such as respecting individual perspectives and rights [7]. It is necessary for all members of society to understand that there are boundaries between freedom of speech and fact-checking. Measures must be implemented to strike a balance between freedom of speech and the verification of information [10]. Freedom of speech is a fundamental principle of a democratic society, but the excessive spread of rumors and false information can have a significant impact on society. Effective fact-checking works towards promoting truth and maintaining the integrity of information, while ensuring respect for individual perspectives and rights. Furthermore, it is necessary to increase the diversity of information sources and establish correction mechanisms for biases in the design and practice of censorship to increase its fairness and rigor [8]. In the application of AI automation technology, the balance between technical neutrality and ethical guidance is an essential factor in maintaining moral credibility [3]. It is necessary to recognize the biases that exist within the algorithms and decision-making processes. Measures should be taken to mitigate these biases and ensure that censorship is conducted in a fair and credible manner. The transparency of AI automation technology should also be emphasized in the information censorship process to enhance its fairness and credibility [9]. The interpretability of algorithm decision-making processes is also crucial for enhancing the credibility and persuasiveness of AI automation technology's algorithm decision-making [9]. Therefore, clear ethical guidelines and technical specifications need to be established to ensure that the balance between technical neutrality and ethical guidance is maintained [6]. It is essential to prevent censorship from being utilized as a tool by authoritarian regimes to restrict freedom of speech [10]. Effective information censorship works towards promoting truth and maintaining the integrity of information, while ensuring respect for individual perspectives and rights. Therefore, it is essential to strike a balance between the freedom of speech and fact-checking in order to accomplish this shared objective [10].

4.   Conclusion

In conclusion, this paper has explored ethical practices concerning AI automated rumor detection technologies, including privacy protection, information censorship transparency, and the balance between freedom of speech. This study proposes relevant laws and regulations, as well as ethical suggestions, aimed at providing guidance for the ethical practices of AI-automated rumor detection technologies. Future studies could further explore the implementation of the proposed legal and ethical guidelines in specific scenarios and industries, such as politics, economics, and healthcare. Additionally, it would be beneficial to investigate the impact of AI automation technology on the social and psychological aspects of individuals and society in order to comprehensively understand the implications of its use. Furthermore, there are still many gaps in existing research on AI automation technology, such as the influence of cultural differences on its application and the potential biases in its decision-making. As such, future research could focus on these gaps to provide a more comprehensive understanding of AI automation technology and its ethical and legal implications. The use of AI automation technology for identifying rumors and misinformation has significant potential in combating the spread of false information and protecting information security. However, attention must also be paid to ethical and legal considerations, such as privacy protection, transparency in information censorship, and the balance between freedom of speech and censorship. The establishment of legal regulations and ethical guidelines can provide a foundation for the ethical and lawful practices of AI automation technology. Future research could enhance the comprehension

of these issues in specific scenarios and industries while also addressing the existing gaps in research. However, a more detailed investigation and discussion of these issues within specific practices are necessary to seek improved solutions.

References

[1] Nissenbaum, H. (1998). Protecting privacy in an information age: The problem of privacy in public. Law and Philosophy, 17(5), 559-596.
[2] Solove, D. J. (2006). A taxonomy of privacy. University of Pennsylvania Law Review, 154(3), 477-564.
[3] Wong, E. (2017). How China Is Fighting Against Fake News? The New York Times.
[4] Zide, L., & Ali, P. (2019). CHAPTER 18: Using data and automation to fight fake news. In Social Commerce and Fake News in the Digital Era (pp. 295-312). IGI Global.
[5] Jia, T., & Liang, X. (2016). Data privacy protection and data sharing: Current debates in China. Telecommunications Policy, 40(9), 817-825.
[6] Floridi, L., & Taddeo, M. (2016). What is data ethics? Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2083), 20160360.
[7] Lavallee, A. (2019). Censorship or misinformation? AI monitoring challenges free speech. Phys.org.
[8] Hoffman, C. P., & Proulx, T. (2018). Reducing bias in social media censorship through the design of impact-based interventions. The Journal of Interactive Technology and Pedagogy, (13).
[9] Resnik, P., & Hardcastle, T. (2018). Transparency at the boundaries of human and machine decision-making. Ethics and Information Technology, 20(1), 7-19.
[10] Tang, D., Zhang, L., Liu, K., & Huang, X. (2008). Twitter topic modeling based on social network structure. Proceedings of the 17th International Conference on World Wide Web, 971-980.