

Understanding the Causes and Solutions of AI-Induced Misinformation Impacting the Decision-Making Behavior of Students

Vo Quoc Huy¹

Abstract

This research aims to investigate the causes and potential solutions for AI-induced misinformation that impacts the decision-making behavior of students. With the increasing prevalence of artificial intelligence (AI) technologies in our daily lives, there is growing concern about the spread of misinformation and its influence on individuals' decision-making processes. This study seeks to explore the underlying factors that contribute to the dissemination of AI-induced misinformation, including algorithm biases, echo chambers, and the lack of critical thinking skills among students. Additionally, the research aims to identify effective strategies and interventions to mitigate the negative effects of AI-induced misinformation on students' decision-making behavior. By understanding the causes and developing potential solutions, this study intends to contribute to the development of informed decision-making practices in the context of AI technologies.

Keywords: education, technology, information, student, behavior, decision-making.

1. Introduction

Artificial intelligence (AI) has become an increasingly pervasive force in our society, revolutionizing various aspects of our daily lives. From personalized recommendations on streaming platforms to automated customer service chatbots, AI technologies have the potential to enhance efficiency and convenience. However, alongside these advancements, concerns have arisen regarding the spread of misinformation facilitated by AI systems. This is particularly impactful in the context of decision-making, where the consequences of misinformation can be significant [9]. The decision-making behavior of students is a critical aspect of their educational journey and future success. However, the prevalence of AI-induced misinformation poses challenges to their ability to make informed choices. AI algorithms, while designed to provide relevant and accurate information, can inadvertently perpetuate biases or promote false narratives. This can occur due to algorithmic biases, echo chambers, and the lack of critical thinking skills among students. Algorithmic biases refer to the inherent biases that AI algorithms can develop due to the data they are trained on. If these algorithms are fed with biased or incomplete data, they may produce information that is skewed or inaccurate. This can mislead students and influence their decision-making processes. Furthermore, echo chambers, which are created when individuals are exposed only to information that aligns with their existing beliefs, can be exacerbated by AI algorithms that personalize content. Students may find themselves trapped in echo chambers, shielded from diverse perspectives and alternative viewpoints. This can limit their exposure to accurate and

¹ Faculty of the Japanese language, FPT University, Vietnam, huyvqsa160101@fpt.edu.vn

well-rounded information, leading to misguided decisions. Additionally, the lack of critical thinking skills among students can make them more susceptible to AI-induced misinformation. Critical thinking involves evaluating information, detecting biases, and questioning the reliability of sources. Insufficient critical thinking skills can hinder students' ability to discern and navigate through misinformation, making them more vulnerable to its influence. Understanding the causes and solutions of AI-induced misinformation impacting the decision-making behavior of students is crucial for ensuring that they can make informed choices in an increasingly AI-driven world. By investigating the underlying factors contributing to misinformation dissemination and exploring potential strategies to mitigate its negative effects, this research aims to provide insights and recommendations that can empower students to navigate the complex landscape of AI-generated information and make well-informed decisions. In the following sections, this study will delve into the causes and effects of AI-induced misinformation, examine the challenges it poses to students' decision-making behavior, and propose potential solutions and interventions to address these issues. By doing so, we hope to contribute to the development of a more informed and discerning society in the face of AI-induced misinformation. On May 16th, the World Health Organization (WHO) issued a warning regarding the cautious use of artificial intelligence (AI) applications in community healthcare, as the data utilized by AI systems can lead to inaccurate or biased results. While WHO recognizes the potential of AI in advancing healthcare, concerns arise about how AI is employed to access medical information, as it is considered a tool to support decision-making and improve diagnoses. In a statement, WHO stated that the data used in AI can be biased, resulting in misleading or incorrect information, and these models can even be exploited to generate misinformation. The organization emphasized the need for risk assessment regarding the use of large language models similar to ChatGPT, to safeguard and promote human well-being and community health. This warning comes at a time when AI applications are rapidly gaining popularity, as they are seen as tools to enhance business quality and transform the way society operates. Prior to this, on May 10th, researchers also urged healthcare experts to issue a global warning about the potential risks that artificial intelligence (AI) technology can pose to human health. AI technology has become even more prevalent last year, following the explosive introduction of ChatGPT, a chatbot with the ability to generate coherent text based on keywords or short prompts. When using AI, we will have to confront and address biases and misinformation regarding gender in LLMs. This is not only a technical challenge but also an ethical responsibility to ensure fairness and integrity in AI. Prevalent forms of biases and misinformation include gender, racial, and cultural biases. Therefore, there is a need to train LLMs in a way that minimizes the impact of biases and misinformation. Additionally, it is important to explore future directions to improve the accuracy and reliability of LLMs [6].

2. Literature review

The term deepfake refers to the product of utilizing AI techniques, specifically through the collaboration of two AI algorithms within a Generative Adversarial Network (GAN). Deepfakes are created by algorithmically generating new data from existing datasets, such as analyzing thousands of pictures of a person and generating a new image that resembles the analyzed images but is not an exact copy of any of them. This technology can be applied to various forms of content, including images, videos, audio, and text. However, the term deepfake is primarily used to describe manipulated audio and video content. The development of AI has expanded the possibilities for manipulating and creating realistic texts, images, audios, and videos, with deepfakes being a notable outcome Walorska, (2020) [1]. An important method to create these models for discriminatory programs is referred to as the "adversarial" system. Adversarial machine learning is the process of creating malicious or misinforming content that can slip past

detection programs. Grover and other AI systems improve their efficiency by generating articles and then using their own detection programs to evaluate the articles' believability. If the created content isn't as convincing, the generators keep reproducing text and learn what is real and what isn't. This ability to "generate" fake articles is what works as a double-edged sword. [2] Another advancement is the creation of "deepfakes," which are doctored or artificially generated videos and photos that can superimpose the physique and face of one person on another to make it seem like they carried out a certain action. Deepfakes can have drastic consequences upon misuse. From propaganda to incite hatred and violence, to maligning public figures with fake speeches and doctored videos, deepfakes can be used to create confusion and can result in a severe loss of public trust and a bad reputation. A prime example of such misuse was the release of a doctored video of Facebook CEO Mark Zuckerberg around the time of his congressional hearing. Although it is difficult to identify whether such media is authentic, the technology to fight this form of AI manipulation is still in the works. [3]

3. Research Methodology

This approach combines both quantitative and qualitative methods to provide a comprehensive understanding of the research topic. Mixed methods research allows researchers to gather and analyze both numerical and non-numerical data, providing a more holistic view of the research question.

4. Research result and discussion

Americans generally express a greater level of concern rather than excitement regarding the increased use of AI. This sentiment suggests that there are apprehensions and reservations about the implications and potential consequences of AI integration in various aspects of daily life. The concerns often revolve around issues such as job displacement, privacy considerations, the possibility of AI surpassing human capabilities, the erosion of human connection, misuse of AI technology, and overreliance on AI systems. While some individuals may still feel a degree of excitement about AI advancements, the prevailing sentiment leans towards caution and vigilance when it comes to the widespread adoption of AI technologies.

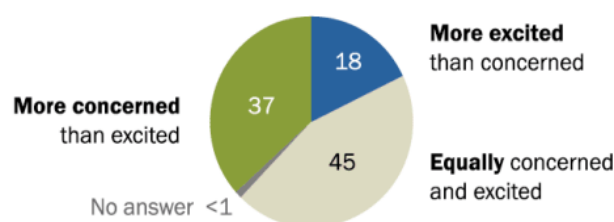


Figure 1. Americans lean toward concern over excitement when it comes to increased of AI in daily life

Source: Pew research centermarch (2022)

In general terms, a greater proportion of Americans express a higher level of worry rather than enthusiasm regarding the growing integration of AI into everyday life. About 45% of adults in the United States indicate they feel both concern and excitement to an equal extent. When asked to elaborate on their concerns about AI, individuals who lean more toward worry than excitement often mention fears about job displacement, privacy issues, the possibility of AI surpassing human capabilities, and concerns about a diminished sense of human connection. Additionally, some express worries about AI being misused or overly relied upon.

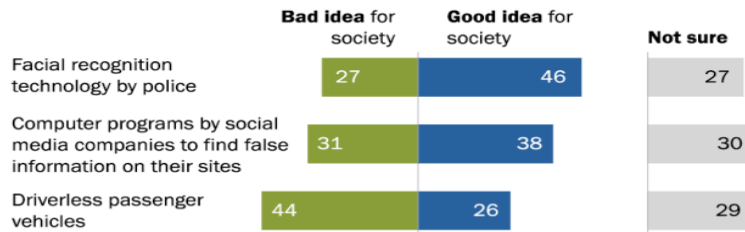


Figure 2. Public views are varied when it comes to three specific AI applications

Source: Pew research center march (2022)

As AI becomes increasingly prevalent in various critical areas of people's lives, such as healthcare, legal services, agriculture, and transportation, Americans are witnessing its widespread adoption. This expansion of AI engenders a mixture of concerns and excitement among the population. AI applications in healthcare, such as diagnostic algorithms and personalized treatment recommendations, can evoke a mix of excitement and concern. Many people are optimistic about the potential of AI to improve medical diagnoses, enhance patient care, and accelerate medical research. However, concerns may arise regarding the privacy and security of personal health data, the potential for bias in algorithms, and the ethical implications of AI making critical healthcare decisions. The deployment of AI in autonomous vehicles can generate a range of opinions. Some individuals are excited about the prospect of increased road safety, reduced congestion, and improved transportation efficiency. [7] Others may express concerns about the reliability and safety of autonomous vehicles, ethical dilemmas in decision-making algorithms, and the potential impact on employment in the transportation industry. The use of AI in surveillance and law enforcement can be a subject of controversy. Supporters argue that AI-powered surveillance systems can enhance public safety and aid in crime prevention and detection. However, concerns may arise around privacy infringement, potential biases in facial recognition algorithms, and the risk of misuse or abuse of AI systems by authorities. Overall, public views on specific AI applications can be diverse, reflecting a complex interplay between perceived benefits and potential risks associated with each application. It is essential to engage in informed and balanced discussions to address concerns and ensure responsible and ethical implementation of AI technologies.

5. Recommendation

Addressing AI-induced misinformation impacting the decision-making behavior of students requires a multi-faceted. Integrate information literacy education into school curricula to equip students with critical thinking skills, source evaluation techniques, and media literacy competencies. This will enable them to identify and assess AI-generated misinformation effectively. Educate students on media literacy, including understanding the biases and limitations of AI algorithms. Teach them how to evaluate sources, fact-check information, and differentiate between credible and unreliable sources. Emphasize the ethical considerations surrounding AI technologies. Teach students about responsible AI development, usage, and potential biases. Encourage them to question the intentions and consequences of AI-generated information. Foster collaboration between educational institutions and technology companies to develop AI systems that prioritize accuracy, transparency, and fairness [5]. Encourage the development of AI tools that help identify and flag misinformation. Provide teachers with training and professional development opportunities on AI literacy and misinformation detection. Equip them with the knowledge and skills to guide students in critically analyzing AI-generated information. Develop and promote tools and platforms that facilitate fact-checking and verification of AI-generated information. Encourage their use to empower students in verifying the accuracy of information they encounter. Encourage collaboration between educators,

researchers, technologists, and policymakers to address AI-induced misinformation. Foster interdisciplinary approaches to develop effective strategies, policies, and educational resources. Support ongoing research and evaluation efforts to understand the evolving nature of AI-induced misinformation and its impact on students' decision-making behavior. This will help inform evidence-based interventions and strategies.

6. Conclusion

The prevalence of AI-induced misinformation in education is a concerning issue that impacts students' decision-making behavior. However, by implementing a range of solutions, we can work towards mitigating its effects and equipping students with the necessary skills to navigate the digital information landscape effectively. [8] Promoting information literacy education, fostering media literacy, encouraging ethical AI use, collaborating with tech companies, providing teacher training, leveraging fact-checking tools, fostering interdisciplinary approaches, and supporting continuous research and evaluation are all key components of addressing this problem. By taking a comprehensive and collaborative approach, we can empower students to critically analyze AI-generated information and make informed decisions in an increasingly AI-driven world.

Conflict of interests

None

Acknowledgement

I am grateful to the broader academic community for their contributions to the field, which have shaped and informed this study. Once again, I express my heartfelt appreciation to all those who have played a part in this research endeavor. Your support and contributions have been vital, and I am truly grateful for your assistance.

References

1. Walorska, A. M. (2020). The Algorithmic Society. *Redesigning Organizations: Concepts for the Connected Society*, 149-160.
2. Bontridder, N., & Poulet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy*, 3, e32.
3. Thompson, N., & Vogelstein, F. (2018). Inside the two years that shook Facebook—and the world. *Wired*. February, 12.
4. Moran, R. E. (2020). Examining switching power: Mark Zuckerberg as a novel networked media mogul. *Information, Communication & Society*, 23(4), 491-506.
5. Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994.
6. Dempere, J., Modugu, K. P., Hesham, A., & Ramasamy, L. (2023). The impact of ChatGPT on higher education. Dempere J, Modugu K, Hesham A and Ramasamy LK (2023) The impact of ChatGPT on higher education. *Front. Educ*, 8, 1206936.
7. Josimovski, S., Ivanovska, L. P., & Dodevski, D. (2023). Understanding the Consumer Dynamics of AI in North Macedonian E-Business. *Economics and Culture*, 20(2), 64-75.
8. Kim, J., Kadkol, S., Solomon, I., Yeh, H., Soh, J. Y., Nguyen, T. M., ... & Ajilore, O. A. (2023). AI Anxiety: A Comprehensive Analysis of Psychological Factors and Interventions. Available at SSRN 4573394.
9. Sebastian, G. (2023). Exploring Ethical Implications of ChatGPT and Other AI Chatbots and Regulation of Disinformation Propagation. Available at SSRN 4461801.