

12-15-2024

Fake news warning system: An exploratory study on using Large Language Model to spot fake news

An Vu-Minh
National Economics University, vuxminhan@gmail.com

Thi Tran
Binghamton University, SUNY, ttran9@binghamton.edu

Duong Nguyen-Quy
National Economics University, quyduong106@gmail.com

Duc-Viet Hoang
University of Science, Vietnam National University, ducviet5138@gmail.com

Sumantra Sarkar
SUNY - Binghamton, ssarkar@binghamton.edu

Follow this and additional works at: https://aisel.aisnet.org/treos_icis2024

Recommended Citation

Vu-Minh, An; Tran, Thi; Nguyen-Quy, Duong; Hoang, Duc-Viet; and Sarkar, Sumantra, "Fake news warning system: An exploratory study on using Large Language Model to spot fake news" (2024). *ICIS 2024 TREOS*. 108.

https://aisel.aisnet.org/treos_icis2024/108

This material is brought to you by the AIS TREO Papers at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2024 TREOS by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Fake news warning system: An exploratory study on using Large Language Model to spot fake news

TREO Talk Paper

An Vu-Minh; vuxminhan@gmail.com

Thi Tran, ttran9@binghamton.edu

Duong Nguyen-Quy, guyduong106@gmail.com

Duc-Viet Hoang, ducviet5138@gmail.com

Sumantra Sarkar, ssarkar@binghamton.edu

In the era of rapid news sharing and social media, fake news are being generated at an unprecedented speed, and spread rapidly once it catches audience attention. This results in detrimental harms done to its audience and needs to be effectively addressed (Tran et al., 2023). Prior research has explored improving fake news detection systems and algorithms (Wang et al., 2024), understanding the psychology behind fake news creation and consumption, and examining how false information spreads. However, to the best of our knowledge, there is no comprehensive solution capable of analyzing a news article with multiple data types holistically—encompassing its title, embedded images, metadata (e.g., hyperlinks, suggested readings, author biography)—while also providing user-friendly explanations for the system’s decisions. Following design science methodology outlined by Hevner et al. (2004), we propose a novel system of fake news detection which integrates state-of-the-art Large Language Model (LLM) agents to enable automatic and in-depth analysis of news articles, offering human-interpretable insights into its conclusions. As the performance, efficiency and cost-effectiveness of LLM is rapidly increasing (Wang et al., 2024), the system will benefit from this development and offer scalability and adaptability to users as a robust tool for combatting the spread of misinformation. We will validate the applicability and effectiveness of our proposed system through two steps. First, we will obtain evaluations on our system, from related domain experts through open ended surveys. Second, we will test the actual performance of the system with extracted accuracy rates when detecting complex fake versus real news articles that have been labelled by reputable fact checkers. We intend to contribute to literature by examining the related contents in information classification. This system will aid existing efforts of designing applications in effectively detecting and mitigating harmful fake news.

References

1. Tran, T., Valecha, R., & Rao, H. R. (2023). Machine and human roles for mitigation of misinformation harms during crises: An activity theory conceptualization and validation. *International Journal of Information Management*, 70, 102627.
2. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 75-105.
3. Wang, B., Ma, J., Lin, H., Yang, Z., Yang, R., Tian, Y., & Chang, Y. (2024, May). Explainable Fake News Detection With Large Language Model via Defense Among Competing Wisdom. In *Proceedings of the ACM on Web Conference 2024* (pp. 2452-2463).