

ARTIFICIAL INTELLIGENCE AS A TOOL FOR COMBATING INFORMATION FALSIFICATION AND MANIPULATION

Vinnitsia National Technical University

Анотація:

У роботі досліджено роль технологій штучного інтелекту у протидії фальсифікації та маніпуляції інформацією в цифровому середовищі. Проаналізовано теоретичні основи застосування інтелектуальних систем для виявлення дезінформації, сучасні методи автоматизованої перевірки фактів, а також проблеми й перспективи розвитку таких технологій. Обґрунтовано, що ефективна боротьба з інформаційними загрозами можлива лише за умови поєднання технічних інструментів, правового регулювання та критичного мислення користувачів.

Ключові слова: штучний інтелект, виявлення дезінформації, фейкові новини, інформаційна маніпуляція, автоматизована перевірка фактів, інформаційна безпека, мультимодальний аналіз, верифікація цифрового контенту.

Abstract:

The paper examines the role of artificial intelligence technologies in countering falsification and manipulation of information in the digital environment. It analyses the theoretical foundations of using intelligent systems to detect disinformation, modern methods of automated fact-checking, as well as the problems and prospects for the development of such technologies. It argues that effective combating of information threats is only possible through a combination of technical tools, legal regulation and critical thinking on the part of users.

Keywords: artificial intelligence, disinformation detection, fake news, information manipulation, automated fact-checking, information security, multimodal analysis, digital media verification.

Introduction

In the modern information society, digital data and media content play a decisive role in shaping public opinion, political processes, and social relations. The rapid growth of information volumes is accompanied by the spread of falsification and manipulation, which pose a threat to societal stability and national security [1]. The dissemination of unreliable information undermines trust in media, provokes social conflicts, and distorts the perception of reality [2].

Under such conditions, the application of intelligent technologies for detecting and preventing disinformation becomes particularly relevant. The object of the study is the process of dissemination of falsified information in the digital environment, while the subject is modern intelligent technologies for its detection and counteraction.

Methodological Approaches to AI-Based Disinformation Detection

Artificial intelligence is defined as a set of methods and technologies that enable computer systems to perform cognitive functions traditionally associated with human intelligence, including data analysis, generalization, and decision-making [3]. In the field of countering disinformation, intelligent systems are used for automated analysis of textual and multimedia content, identification of information dissemination patterns, and verification of the credibility of statements. Automated fact-checking involves formulating a claim, searching for supporting evidence, and evaluating the consistency of the information with reliable sources, while the most challenging stage is the formation of a high-quality evidence base.

Modern artificial intelligence technologies are already widely applied to detect fake news, deepfake content, bot activity, and manipulative information campaigns in social networks. An important direction of development is multimodal analysis, which combines processing of textual, audio, and video data and enables detection of inconsistencies between different types of information [4]. The analysis of information dissemination networks also represents an effective tool, allowing the identification of key nodes and coordinated disinformation campaigns [5]. Technologies for verifying content authenticity, including digital

watermarking and data provenance verification methods, play a significant role in enhancing transparency in the information environment.

Despite considerable progress, the application of artificial intelligence in combating information manipulation is associated with several challenges. These include the lack of high-quality training data, difficulties in adapting models to new forms of disinformation, insufficient algorithmic transparency, and limited integration of intelligent systems into information platforms. In addition to technical limitations, legal and ethical challenges related to the regulation of artificial intelligence and liability for manipulative content dissemination remain significant [6]. Prospects for further development of intelligent systems are associated with improving model explainability, enhancing multimodal analysis, and integrating artificial intelligence technologies into national information security frameworks [7].

Conclusions

The conducted analysis demonstrates that artificial intelligence is an effective tool for counteracting information falsification and manipulation. Modern intelligent technologies enable automated fact-checking, detection of coordinated information campaigns, and improvement of the credibility of digital content. At the same time, their effectiveness depends on the integration of technological capabilities with legal regulation and the development of users' critical thinking. The synergy between human analytical capacity and intelligent technologies forms the foundation for sustainable resistance to information threats and contributes to the development of a responsible information society.

REFERENCES

1. Russell, S., Norvig, P. Artificial Intelligence: A Modern Approach. 4th ed. – Pearson, 2022. – [Електронний ресурс]. – Режим доступу: http://lib.ysu.am/disciplines_bk/efdd4d1d4c2087fe1cbe03d9ced67f34.pdf
2. Romero, O. J., Zimmerman, J., Steinfeld, A., Tomasic, A. Synergistic Integration of Large Language Models and Cognitive Architectures for Robust AI: An Exploratory Analysis. – 2023. – [Електронний ресурс]. – Режим доступу: <https://arxiv.org/abs/2308.09830>
3. Goodfellow, I., Bengio, Y., Courville, A. Deep Learning: Foundations and Trends. – MIT Press, 2023. – [Електронний ресурс]. – Режим доступу: <https://pzs.dstu.dp.ua/DataMining/bibl/Deep%20Learning.pdf>
4. Garcez, A. d'Avila, Lamb, L. C. Neurosymbolic AI: The 3rd Wave // Artificial Intelligence Review. – 2023. – Т. 56, № 11. – С. 12387–12406. – DOI: <https://doi.org/10.1007/s10462-023-10448-w>
5. Muravska, Y. Legal regulation of artificial intelligence in Ukraine and worldwide // Actual Problems of Law. – 2024. – [Електронний ресурс]. – Режим доступу: <https://appj.wunu.edu.ua/index.php/apl/article/view/1788>
6. Aïmeur, E., Amri, S., Brassard, G. Fake news, disinformation and misinformation in social media: a review // Social Network Analysis and Mining. – 2023. – Т. 13, № 1. – С. 30. – [Електронний ресурс]. – DOI: <https://doi.org/10.1007/s13278-023-01028-5>
7. Altuncu, E., Franqueira, V. N. L., Li, S. Deepfake: definitions, performance metrics and standards, datasets, and a meta-review // Frontiers in Big Data. – 2024. – Т. 7. – Стаття 1400024. – [Електронний ресурс]. – DOI: <https://doi.org/10.3389/fdata.2024.1400024>

Yankivska Bohdana Anatoliivna — student of group 2CS-24b, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: bjankivska@gmail.com

Scientific Supervisor: **Melekh Anna Vasylivna** – English language instructor at the Department of Foreign Languages, Vinnytsia National Technical University, email: a.melekh@vntu.edu.ua.

Янківська Богдана Анатоліївна — студентка групи 2КН-24б, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: bjankivska@gmail.com

Науковий керівник: **Мелех Анна Василівна** – викладач англійської мови, кафедра іноземних мов, Вінницький національний технічний університет, email: a.melekh@vntu.edu.ua.