Applied Data Science and Analysis
Vol.2023, **pp**. 108–109
DOI: https://doi.org/10.58496/ADSA/2023/009 ; ISSN: 3005-317X
https://mesopotamian.press/journals/index.php/ADSA

Letter to Editor

The Considerations of Trustworthy AI Components in Generative AI; A Letter to Editor

A. S. Albahri [1,*], Yahya Layth Khaleel [2], Mustafa Abdulfattah Habeeb [2]

[1] *Faculty of Computing and Meta-Technology (FKMT), Universiti Pendidikan Sultan Idris (UPSI), Perak, Malaysia*

[2] *Faculty of Mechanical Engineering and Informatics, University of Miskolc, H-3515 Miskolc, Hungary*

Dear Editor: In navigating the ever-expanding realms of Artificial Intelligence (AI) across diverse domains, particularly within the purviews of Generative AI, it becomes imperative to delve into the intricate considerations of trustworthy AI components [1]. This letter aims to underscore the salient aspects of this discussion with a specific focus on the thematic areas championed by the journal.

Much like the uncertainties accompanying unprecedented events, generative AI introduces complexities and potential risks that must be addressed within the context of data mining, knowledge discovery, and machine learning [2]. Recently, the authors in [1] identified the components of trustworthy AI, featuring seven main components, each with subcomponents: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination, and fairness, societal and environmental well-being, and finally, accountability.

In lieu of explicit examples, we can draw parallels between the potential stress and anxiety analogous to uncertain situations and users' concerns about the reliability, transparency, and ethical considerations of generative AI systems. This pervasive lack of clarity may lead to apprehension and, in certain instances, avoidance of AI applications, particularly pertinent to the realms of big data analytics, predictive analytics, and forecasting [3]. The ethical considerations, potential biases, and consequences of AI-generated content also demand thorough evaluation within the realms of Natural Language Processing (NLP), computer vision, and image analysis [4]. As we navigate these intricacies, it is vital to remain cognizant of the multifaceted challenges faced in different applications and scenarios, especially within Social Network Analysis.

To address these concerns, we must enhance awareness regarding trustworthy AI components in Generative AI across various thematic domains. This involves educating users, developers, and decision-makers on ethical considerations, potential biases, and mechanisms ensuring transparency and accountability in AI systems [5][6]. Additionally, other theories such as Multi-Criteria Decision-Making (MCDM) and fuzzy sets can be applied and/or integrated with AI to achieve trustworthiness in decision-making [7]. For example, in the problem of datasets without labeling (target), MCDM can assign and process the appropriate target for these cases in the dataset based on the formulation of several models using subjective judgment from experts [8]. In these contexts, machine learning algorithms can be applied to the processed dataset. Other examples can be found in [9][10]. Then, Generative AI can play a pivotal role in revolutionizing decision-making processes.

Such an approach aligns seamlessly with the broader focus of the journal on ethical and privacy issues in data science. In the spirit of promoting user-centric and responsible AI models, this letter calls for a comprehensive and nuanced understanding of trustworthy AI components in generative AI, with a specific emphasis on the thematic areas championed by the journal. By assimilating lessons from various domains, we contribute to the responsible development and deployment of generative AI, fostering trust, and mitigating potential risks in data engineering and data infrastructure.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

Funding

The authors receive no funding for this work.

Acknowledgment

None.

*Corresponding author. Email: ahmed.bahri1978@gmail.com

References

[1] A. S. Albahri et al., “A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion,” Inf. Fusion, vol. 96, pp. 156–191, Mar. 2023, doi: 10.1016/j.inffus.2023.03.008.

[2] L. Alzubaidi et al., “Towards Risk-Free Trustworthy Artificial Intelligence: Significance and Requirements,” Int. J. Intell. Syst., vol. 2023, p. 4459198, 2023, doi: 10.1155/2023/4459198.

[3] K. B. Ooi et al., “The Potential of Generative Artificial Intelligence Across Disciplines: Perspectives and Future Directions,” J. Comput. Inf. Syst., pp. 1–32, 2023, doi: 10.1080/08874417.2023.2261010.

[4] P. P. Ray, “ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope,” Internet Things Cyber-Physical Syst., vol. 3, pp. 121–154, 2023, doi: 10.1016/j.iotcps.2023.04.003.

[5] WHO, “Ethics and Governance of Artificial Intelligence for Health: WHO guidance,” World Heal. Organ., pp. 1–148, 2021.

[6] C. Adams, P. Pente, G. Lemermeyer, and G. Rockwell, “Ethical principles for artificial intelligence in K-12 education,” Comput. Educ. Artif. Intell., vol. 4, no. 4, pp. 4221–4241, 2023, doi: 10.1016/j.caeai.2023.100131.

[7] A. S. Albahri et al., “A Systematic Review of Using Deep Learning Technology in the Steady-State Visually Evoked Potential-Based Brain-Computer Interface Applications: Current Trends and Future Trust Methodology,” Int. J. Telemed. Appl., vol. 2023, 2023, doi: 10.1155/2023/7741735.

[8] M. Baydaş and D. Pamučar, “Determining Objective Characteristics of MCDM Methods under Uncertainty: An Exploration Study with Financial Data,” Mathematics, vol. 10, no. 7, p. 1115, 2022, doi: 10.3390/math10071115.

[9] S. S. Joudar, A. S. Albahri, and R. A. Hamid, “Intelligent triage method for early diagnosis autism spectrum disorder (ASD) based on integrated fuzzy multi-criteria decision-making methods,” Informatics Med. Unlocked, vol. 36, p. 101131, 2023, doi: 10.1016/j.imu.2022.101131.

[10] N. S. Baqer et al., “Indoor air quality pollutants predicting approach using unified labelling process-based multi-criteria decision making and machine learning techniques,” Telecommun. Syst., vol. 81, no. 4, pp. 591–613, Dec. 2022, doi: 10.1007/s11235-022-00959-2.