frontiers | Frontiers in Computer Science

TYPE Editorial
PUBLISHED 15 August 2023
DOI 10.3389/fcomp.2023.1265902

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Kostas Karpouzis,
Panteion University, Greece

*CORRESPONDENCE
Renwen Zhang
r.zhang@nus.edu.sg

RECEIVED 24 July 2023
ACCEPTED 03 August 2023
PUBLISHED 15 August 2023

CITATION
Zhang R, Zhang Z, Wang D and Liu Z (2023) Editorial: Responsible AI in healthcare: opportunities, challenges, and best practices. *Front. Comput. Sci.* 5:1265902. doi: 10.3389/fcomp.2023.1265902

Editorial: Responsible AI in healthcare: opportunities, challenges, and best practices

Renwen Zhang[1]*, Zhan Zhang[2], Dakuo Wang[3] and Ziwei Liu[4]

[1]Department of Communications and New Media, National University of Singapore, Singapore, Singapore, [2]School of Computer Science and Information Systems, Pace University, New York City, NY, United States, [3]Khoury College of Computer Sciences and the College of Arts, Media and Design, Northeastern University, Boston, MA, United States, [4]School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore

KEYWORDS

healthcare, Clinical Decision Support Systems (CDSS), Artificial Intelligence, ethics, bias

Editorial on the Research Topic
Responsible AI in healthcare: opportunities, challenges, and best practices

As Artificial Intelligence (AI) makes its way into healthcare, it promises to revolutionize clinical decision-making processes. AI-powered Clinical Decision Support Systems (AI-CDSS) offer the potential to augment clinicians' decision-making abilities, improve diagnosis accuracy, and personalize treatment plans (Magrabi et al., 2019; Montani and Striani, 2019; Giordano et al., 2021). However, with this transformative potential come significant ethical challenges, such as issues of bias, transparency, accountability, and privacy (Keskinbora, 2019; Wang et al., 2021). These challenges have accelerated research on responsible AI, which seeks to ensure that AI systems are developed and deployed in a manner that is ethical, fair, transparent, accountable, and beneficial to all users (Dignum, 2019; Floridi et al., 2021; Floridi and Cowls, 2022). These ethical aspects gain heightened significance in high-stakes domains such as healthcare. This Research Topics features four articles that delve into different aspects of responsible AI in healthcare, including data biases, transparency in uncertainty communication, integration of AI into healthcare, and evaluation of AI-CDSS. In this editorial, we introduce these four articles and provide a brief overview of these critical areas, highlighting the necessity to address these issues to ensure responsible and effective use of AI in healthcare.

Data and algorithmic bias

Bias, whether in data or algorithms, is a cardinal ethical concern in AI-CDSS. Data bias arises when data used to train the AI models are not representative of the entire patient population. This can lead to erroneous conclusions, misdiagnoses, and inappropriate treatment recommendations, disproportionately affecting underrepresented populations (Ganju et al., 2020). Model bias occurs when AI algorithms inherently favor certain outcomes or predictions over others due to their mathematical constructs. Such biases can compromise the fairness and effectiveness of AI-powered CDSS and perpetuate health disparities.

Yogarajan et al. investigates data and algorithmic bias in electronic health records (EHRs) in New Zealand. In response to the need to develop socially responsible and fair AI in healthcare for the New Zealand population, especially indigenous populations, the authors analyzed health data collected by clinicians to examine biases regarding data collection and model development using established techniques and fairness metrics. This study showed evident bias in the data and machine learning models employed in this study to predict preventable harm. The sources of bias may include missing data, small sample size and commonly available pre-trained embeddings to represent text data. This research underscores the crucial need to develop fair, socially responsible machine learning algorithms to enhance healthcare for underrepresented and indigenous populations, such as New Zealand's Māori.

Transparency and communication of uncertainty

AI models are often regarded as "black boxes" due to their complex and opaque decision-making processes. This opacity becomes ethically problematic when AI-CDSS are employed in healthcare. Clinicians and patients must understand the AI's predictions, including the inherent uncertainties, to make informed decisions. A lack of understanding of the inner workings of AI predictions also remains a key barrier to their responsible adoption in clinical workflows (Tonekaboni et al., 2019). However, AI models often lack transparency in communicating these uncertainties, which can impede trust and appropriate use of these systems.

Prabhudesai et al. address the challenge of quantifying and communicating uncertainty in Deep Neural Networks (DNNs) used for medical image segmentation, specifically in brain tumor segmentation. While DNNs provide accurate predictions, they lack transparency in conveying uncertainty, which can lead to false impressions of reliability and potential harm in patient care. The authors propose a computationally-efficient approach called partially Bayesian neural networks (pBNN), which performs Bayesian inference on a strategically selected layer of the DNN to approximate uncertainty. They demonstrate the effectiveness of pBNN in capturing uncertainty for a large U-Net model and showcase its potential for clinicians to interpret and understand the model's behavior. The methodology proposed by the authors holds promise of empowering clinicians in their interaction with AI-based CDSS and facilitating safer and more responsible integration of AI-CDSS in clinical workflows.

Evaluation of AI-CDSS

A substantial body of research has focused on developing innovative algorithms to enhance the technical performance of AI-CDSS (Alloghani et al., 2019; Barragán-Montero et al., 2021). However, relying solely on technological advancements is inadequate to ensure the successful implementation and user adoption of AI-CDSS. Recent studies have emphasized the significance of investigating human, social, and contextual factors that play a crucial role in the adoption of AI-CDSS (He et al., 2019; Schoonderwoerd et al., 2021). Consequently, there is a growing interest in the human-centered design of AI-CDSS and the exploration of fairness and transparency in AI. Therefore, it is imperative to synthesize the knowledge and experiences reported in this research area to shed light on future investigations.

Wang et al.'s systematic review effectively addresses this research gap. Their article provides valuable insights into the methodologies and tools employed for evaluating AI-CDSS, which can greatly benefit researchers. Furthermore, the review identifies various challenges associated with implementing AI-CDSS interventions, including workflow misalignment, attitudinal, informational, and environmental barriers, as well as usability issues. These challenges underscore the importance of examining and addressing sociotechnical obstacles in the implementation of AI-CDSS. The article also discusses several future research directions and design implications that can guide upcoming studies.

Integration of AI in healthcare

New system implementation in healthcare institutions is often accompanied by a change in clinical workflow and organizational culture (Zhang et al., 2019). Despite numerous efforts in advancing clinical decision support tools, most of these tools have failed in practice. Empirical research has diagnosed poor contextual fit as the cause, such as a lack of consideration of clinicians' workflow and the collaborative nature of clinical work (Wears and Berg, 2005). Thus, foundational research is needed to understand and improve expert work in an age of AI-assisted work, by integrating the richness of context and redefining the role of AI technology in clinical practice.

The paper by Ulloa et al. addresses the invisible labor involved in the integration of medical AI tools in healthcare. Through three case studies, the authors identify four types of labor: data labeling with clinical expertise, identifying algorithmic errors, translating output to patient care decisions, and fostering awareness of AI use. The authors highlight the need for standardized methodologies, reducing clinician burden, formalizing translation processes, and establishing social transparency to foster the adoption and integration of medical AI tools. Integration into existing workflows, usability, documentation, and ethical considerations are also crucial. The authors call for improved documentation of labor, workflows, and team structures to inform future implementations and prevent replicated efforts. They highlight the significance of recognizing and valuing the invisible labor involved in AI development and its impact on system implementation and society as a whole. The paper contributes to understanding the challenges and requirements associated with implementing AI in healthcare, emphasizing the need for a comprehensive approach that considers the labor and sociotechnical aspects to ensure successful and ethical adoption of medical AI tools.

Conclusion

As AI-powered CDSS herald a new era in healthcare, they bring along significant ethical issues that require urgent attention. Bias in data and models, lack of transparency, challenges in integration, and the complexities in evaluation present critical hurdles in harnessing the full potential of AI in healthcare. The four articles in this Research Topic have attempted to address these issues. Addressing these issues is crucial to ensuring that AI-powered CDSS are used responsibly and ethically, upholding the principles of fairness, transparency, and patient-centered care. As we continue to embrace AI's promise, it is essential that we also confront its ethical and contextual complexities, crafting an AI-infused future that is not just technologically advanced, but also user-centered and ethically sound.

Author contributions

RZ: Writing—original draft, Writing—review and editing. ZZ: Writing—review and editing. DW: Writing—review and editing. ZL: Writing—review and editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alloghani, M., Al-Jumeily, D., Aljaaf, A. J., Khalaf, M., Mustafina, M., and Tan, S. (2019). "The application of artificial intelligence technology in healthcare: a systematic review," in *International Conference on Applied Computing to Support Industry: Innovation and Technology*. (Cham: Springer), 248–261.

Barragán-Montero, A., Javaid, U., Valdés, G., Nguyen, D., Desbordes, P., Macq, B., et al. (2021). Artificial intelligence and machine learning for medical imaging: a technology review. *Physica Medica* 83, 242–256. doi: 10.1016/j.ejmp.2021.04.016

Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Cham: Springer International Publishing.

Floridi, L., and Cowls, J. (2022). "A unified framework of five principles for AI in society," in *Machine Learning and the City* (Hoboken, NJ: John Wiley & Sons, Ltd), 535–545. doi: 10.1002/9781119815075.ch45

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., et al. (2021). "An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," in *Ethics, Governance, and Policies in Artificial Intelligence,* Floridi, L. (ed.). (Cham: Springer International Publishing), 19–39. doi: 10.1007/978-3-030-81907-1_3

Ganju, K. K., Atasoy, H., McCullough, J., and Greenwood, B. (2020). The role of decision support systems in attenuating racial biases in healthcare delivery. *Manage. Sci.* 66, 5171–5181. doi: 10.1287/mnsc.2020.3698

Giordano, C., Brennan, M., Mohamed, B., Rashidi, P., Modave, F., and Tighe, P. (2021). Accessing artificial intelligence for clinical decision-making. *Front. Digital Health* 3, 645232. doi: 10.3389/fdgth.2021.645232

He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., and Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nat. Med.* 25, 30–36. doi: 10.1038/s41591-018-0307-0

Keskinbora, K. H. (2019). Medical ethics considerations on artificial intelligence. *J. Clini. Neurosci.* 64, 277–282. doi: 10.1016/j.jocn.2019.03.001

Magrabi, F., Ammenwerth, E., McNair, J. B., De Keizer, N. F., Hyppönen, H., Nykänen, P., et al. (2019). Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. *Yearb. Med. Inform.* 28, 128–134. doi: 10.1055/s-0039-1677903

Montani, S., and Striani, M. (2019). Artificial intelligence in clinical decision support: a focused literature survey. *Yearb. Med. Inform.* 28, 120–127. doi: 10.1055/s-0039-1677911

Schoonderwoerd, T. A., Jorritsma, W., Neerincx, M. A., and Van Den Bosch, K. (2021). Human-centered XAI: Developing design patterns for explanations of Clinical Decision Support Systems. *Int. J. Hum. Comput. Stud.* 154, 102684. doi: 10.1016/j.ijhcs.2021.102684

Tonekaboni, S., Joshi, S., McCradden, M. D., and Goldenberg, A. (2019). "What clinicians want: contextualizing explainable machine learning for clinical end use," in *Machine Learning for Healthcare Conference* (New York City, NY: PMLR), 359–380.

Wang, D., Wang, L., Zhang, Z., Wang, D., Zhu, H., Gao, Y., et al. (2021). "'Brilliant AI doctor' in rural clinics: Challenges in AI-powered clinical decision support system deployment," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama: Association for Computing Machinery), 1–18. doi: 10.1145/3411764.3445432

Wears, R. L., and Berg, M. (2005). Computer technology and clinical work: still waiting for Godot. *JAMA* 293, 1261–1263. doi: 10.1001/jama.293.10.1261

Zhang, R., Burgess, E. R., Reddy, M. C., Rothrock, N. E., Bhatt, S., Rasmussen, L. V., et al. (2019). Provider perspectives on the integration of patient-reported outcomes in an electronic health record. *JAMIA Open* 2, 73–80. doi: 10.1093/jamiaopen/ooz001