BMJ Health & Care Informatics

Call for the responsible artificial intelligence in the healthcare

Umashankar Upadhyay ,[1,2,3] Anton Gradisek,[4] Usman Iqbal,[5,6,7] Eshita Dhar,[1,2] Yu-Chuan Li ,[8] Shabbir Syed-Abdul [1,2]

To cite: Upadhyay U, Gradisek A, Iqbal U, *et al*. Call for the responsible artificial intelligence in the healthcare. *BMJ Health Care Inform* 2023;**30**:e100920. doi:10.1136/bmjhci-2023-100920

Received 06 October 2023
Accepted 12 December 2023

For numbered affiliations see end of article.

Correspondence to
Dr Shabbir Syed-Abdul; drshabbir@tmu.edu.tw

ABSTRACT

The integration of artificial intelligence (AI) into healthcare is progressively becoming pivotal, especially with its potential to enhance patient care and operational workflows. This paper navigates through the complexities and potentials of AI in healthcare, emphasising the necessity of explainability, trustworthiness, usability, transparency and fairness in developing and implementing AI models. It underscores the 'black box' challenge, highlighting the gap between algorithmic outputs and human interpretability, and articulates the pivotal role of explainable AI in enhancing the transparency and accountability of AI applications in healthcare. The discourse extends to ethical considerations, exploring the potential biases and ethical dilemmas that may arise in AI application, with a keen focus on ensuring equitable and ethical AI use across diverse global regions. Furthermore, the paper explores the concept of responsible AI in healthcare, advocating for a balanced approach that leverages AI's capabilities for enhanced healthcare delivery and ensures ethical, transparent and accountable use of technology, particularly in clinical decision-making and patient care.

INTRODUCTION

Advancements in computing power have heightened the prominence of artificial intelligence (AI) in healthcare, thanks to its vast array of applications.[1] From handling patient questions to assisting in surgeries and pushing forward pharmaceutical innovations, AI is offering notable advantages to both patients and the overall healthcare infrastructure.[1] According to Statista, the AI healthcare market, which was valued at US$11 billion in 2021, is forecasted to reach a staggering US$187 billion by 2030.[2] The neural networks and derived deep learning-based AI algorithms lack clarity and transparency, leading clinicians to hesitate or feel uncertain when making prognosis and diagnosis decisions. The key question is how a clinician can convince by technologist with the evidence of the responses. The gap between the AI algorithms and understanding of humans is known as the 'black box' transparency. It is challenging to decide how users can trust that the outcomes of algorithms are correct and appropriate in respect to the analysis, in view of a particular medical situation. There is a common agreement that it is important to consider the explainability of AI seriously, ensuring user's trust and confidence.[2] Although the research and development of AI in healthcare has been ongoing for several decades, the current situation of the AI hype is way different than previous studies.[3] After 2018, there was a sudden increase in the domain of explainable AI (EXAI), with 600 articles per year. However, there were only 92 devices approved by the Food and Drug Administration (FDA) in 2022.[4] The development of deep learning technologies has an impact on the way we look at AI tools and is one of the reasons behind the excitement surrounding AI applications.[1] Healthcare costs are skyrocketing, and the development of costly new therapies contributes to the development of new AI technologies.[5] AI promises to alleviate the impact of this development by the improvement of healthcare and making it more cost-effective.[6]

AI comes with a novel element to healthcare and its relationships.[7] But revolutions rarely come without side effects. There are various concerns related to the use of AI in healthcare. Due to the massive use and advancement of AI technologies worldwide, questions have arisen regarding its impact on societal and individual issues.[8] Over the last 5 years, private companies, research institutions and public sector organisations have issued ethical AI principles and guidelines. It needs to be stressed that AI should be used appropriately to ensure ethics, transparency and accountability. Despite an apparent consensus that AI should be 'ethical', there is a disagreement about what constitutes 'ethical AI', as well as which ethical requirements, technical standards and best practices are required for its realisation. Calls for regulations and policies are getting louder, which has led to the introduction of the concept of responsible AI.[9]

In clinical practice, AI already often plays a major role in clinical decision support systems to assist the clinicians to make better and faster decisions in the diagnosis and treatment of the patients.[10] These applications improve the quality of life of patients and healthcare providers including clinicians. Currently, the healthcare industry has aligned its operations with the vision of Healthcare 4.0, but it is soon approaching the dawn of another paradigm shift, termed Healthcare 5.0. This upcoming shift in healthcare will be more analytical and involve smart controls, virtual reality, and three-dimensional modelling.[11] Thus, healthcare will be smarter, more personalised and dynamic, which includes more reason-based analytics with innovative business solutions. Advanced 5G network and IoT-based sensors integrated with mobile communications will make the healthcare technologies easier to deliver the remote communities.[11] The development in healthcare promises to produce vast amounts of medical data including electronic patient records, images, as well as wearable and other sensor data. AI algorithms such as neural networks will perform complex analytics to process all the healthcare data to enable accurate disease prediction, detection and remote healthcare treatment.[12]

The incorporation of AI support in general practice is increasingly essential, particularly under time pressures that can lead to diagnostic oversights. In the UK, instances of missed measles cases and misdiagnosed appendicitis during winter months have highlighted the need for improved diagnostic precision. AI systems, as evidenced by studies such as Miotto *et al*[13] and Rajkomar *et al*,[14] offer enhanced diagnostic accuracy by identifying subtle patterns and recognising specific symptoms that may be overlooked by human practitioners.

In this article, we adopt a multidisciplinary view of the major healthcare AI challenges: explainability, trustworthiness, transparency and usability. We refer to the challenges of AI in healthcare throughout the manuscript and provide the necessary context for understanding.

CORE CONCEPTS

Explainability of AI has become one of the most debated topics, with implications that extend far beyond technical aspects. AI already outperforms humans in several analytics.[15] While neural networks and associated deep learning approaches are popular due to their powerful performance, they typically act as 'black boxes', not providing users with insights into why a certain decision has been made. Compare this to a simple machine learning model, such as a decision tree, where reconstructing the path from the starting parameters to a decision is straightforward. There are numerous tools and approaches in AI that offer explainability.[9] The lack of explainability has been criticised in the medical domain, while legal and ethical uncertainties may impede progress and prevent AI from fulfilling its potential to improve the lives of patients and healthcare professionals.[16] This all led to the development of the concept of EXAI. EXAI, based on feature engineering, enables the interpretability and explainability of AI algorithms.[16] It is applied to different decision support systems to ensure trustworthy analytics and is used to manage large datasets, helping reduce bias and aiding in disease classification or segmentation.[17]

Trustworthiness of AI systems is crucial for their acceptance and effective use in various applications. Users should have trust and confidence in the system's output, as highlighted in the research by Cutillo *et al*[18] and Laato *et al*[19] In other words, the trust of the users in AI-driven decisions is contingent on the system being perceived as valid and reliable.

Usability refers to the user's ability to understand and use an AI model effectively. This encompasses comprehending the system's goals, scalability and recognising its limitations. Cutillo *et al*[18] underline that usability is key to ensuring that users can harness the potential of AI models. For instance, in business settings, understanding the objectives and limitations of AI-driven analytics tools is essential for users to make informed decisions and leverage the technology effectively.

Transparency and fairness are essential for building trust in AI systems. Users need to understand the system's mechanics and the influence of different inputs on its outcomes. Studies by Cutillo *et al*[18] and Laato *et al*[19] highlight the significance of transparent AI models. When users have access to information about the model's inner workings, they are more likely to trust its decisions. Moreover, transparent models are critical for ensuring fairness and preventing bias in AI systems, as they allow users to closely examine and understand the decision-making process.[20]

In view of ethics in medicine, explainability improves the trustworthiness of the AI applications. Perhaps the strongest benefit comes from uncovering potential biases in the AI models.[19] As these models heavily rely on training data, they can reflect sampling bias, such as the over-representation of a specific demographic that does not generalise to the target population.[21] This can be harmful to under-represented and vulnerable groups. Other types of biases to mention include exclusion bias, where features or instances that could explain trends in the data are omitted, and prejudice bias, where stereotypes directly or indirectly influence the dataset. Considering explainability in the development of AI models for medicine directly benefits the discussion about responsibility in their use, as it offers safety-checks along the road. Furthermore, explainable methods often provide novel insight into the dataset and can be used for knowledge discovery.[22] The lack of scientific knowledge may lead to unintended consequences in emergency responses, thus remaining a fundamental research gap and obstructing the creation of new knowledge.

Moreover, the contention that AI models encode human experience introduces the challenge of inherent biases, as discussed by Buolamwini and Gebru.[23] They highlight how biased training data can result in discriminatory outcomes, emphasising the importance of addressing

biases at the model development stage to ensure fairness. The limitation of AI models by the experiences of their developers, as argued by Mittelstadt *et al*,[24] raises concerns about the potential perpetuation of existing biases and the lack of diversity in perspectives during the model creation process.

The argument that dependence on human expertise limits innovation potential in AI models is well founded, particularly in domains such as cancer progression. The reliance on human-guided categorisation, such as tissue type, can indeed restrict the development of models with a more profound causal understanding. The call for innovation in cancer modelling is echoed by Hoadley *et al*,[25] who advocate for integrative approaches that go beyond traditional classifications and consider diverse data types to enhance the accuracy and insightfulness of AI models.

IMPLEMENTATION CHALLENGES ACROSS THE GLOBE

The evaluation of AI in healthcare presents a complex landscape, particularly when considering its implementation across different global regions. While the potential benefits of AI in healthcare are substantial, varying socioeconomic conditions, healthcare infrastructures, regulatory frameworks and cultural factors can significantly impact the adoption and effectiveness of AI technologies.

In high-income regions, such as North America and Western Europe, where well-established healthcare systems exist, the primary implementation challenge lies in ensuring the seamless integration of AI tools with existing workflows and data systems while adhering to stringent privacy regulations. In contrast, low-income and middle-income regions, such as parts of Africa and Southeast Asia, face challenges related to resource constraints, including limited access to quality data and healthcare professionals. Additionally, ensuring that AI algorithms are culturally and linguistically appropriate is crucial.

Disparities in healthcare access and resources between urban and rural areas can affect the equitable implementation of AI in healthcare. It is important to note that these issues can vary within regions and are subject to change over time. Successful AI implementation in healthcare requires a deep understanding of the local context, collaboration with stakeholders and a tailored approach to address region-specific challenges.

Moreover, cultural and ethical considerations may differ, influencing the acceptance and adoption of AI-driven healthcare solutions. Bridging these disparities in AI healthcare implementation demands a multifaceted approach that encompasses not only technological advancements but also policy harmonisation, capacity building and global collaboration to realise the full potential of AI in healthcare across diverse global regions. Careful consideration of these factors is essential to ensure compliance with local regulations, respect for cultural norms and the development of adaptable solutions.

CONCLUSION

As healthcare increasingly integrates AI into its core operations, the call for responsible AI becomes not just advisable, but imperative. The delicate nature of healthcare decisions, combined with the vast potential of AI, mandates an ethical, transparent and accountable approach. By emphasising responsibility in AI's deployment, we safeguard patient trust, ensure data privacy and uphold the time-honoured principles of medical ethics. The fusion of technology and healthcare holds vast promise, but only if we navigate its intricacies with diligence and conscientiousness. Hence, the drive towards AI in healthcare must be paralleled with an unwavering commitment to its responsible use.

Author affiliations
[1]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, New Taipei City, Taiwan
[2]International Center for Health Information Technology (ICHIT), College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan
[3]Faculty of Applied Sciences and Biotechnology, Shoolini University of Biotechnology and Management Sciences, Solan, India
[4]Department of Intelligent Systems, Jozef Stefan Institute, Ljubljana, Slovenia
[5]Department of Health, Health ICT, Hobart, Tasmania, Australia
[6]School of Population Health, Faculty of Medicine and Health, University of New South Wales, New South Wales, Sydney, Australia
[7]Global Health and Health Security Department, College of Public Health, Taipei Medical University, Taipei, Taiwan
[8]Graduate Institute of Biomedical Informatics, College of Medical Science & Technology, Taipei Medical University, Taipei, Taiwan

Correction notice This article has been corrected since it was published. The affiliations for the author 'Usman Iqbal' has been corrected.

Contributors Conceptualisation: SS-A and Y-CL; writing—original draft preparation: UU, AG and UI; writing—review and editing, SS-A and Y-CL; visualisation, UU and ED; supervision, SS-A.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Commissioned; externally peer reviewed.

ORCID iDs
Umashankar Upadhyay http://orcid.org/0000-0002-2581-8007
Yu-Chuan Li http://orcid.org/0000-0001-6497-4232
Shabbir Syed-Abdul http://orcid.org/0000-0002-0412-767X

REFERENCES

1 Bohr A, Memarzadeh K. n.d. Chapter 2 - the rise of artificial intelligence in Healthcare applications.
2 Education I. 2023. Available: https://www.ibm.com/blog/the-benefits-of-ai-in-healthcare/
3 Lee K-F. *AI Superpowers: China, Silicon Valley, and the New World Order*. Harcourt, Bosto: Houghton Mifflin, 2018.
4 Chaddad A, Peng J, Xu J, *et al*. Survey of explainable AI techniques in healthcare. *Sensors (Basel)* 2023;23:634.
5 Higgins D, Madai VI. From bit to bedside: a practical framework for artificial intelligence product development in healthcare. Adv

Intell Syst 2020;2:10. 10.1002/aisy.202000052 Available: https://onlinelibrary.wiley.com/toc/26404567/2/10
6 Verma A, Bhattacharya P, Zuhair M, *et al*. Vacochain: blockchain-based 5G-assisted UAV vaccine distribution scheme for future pandemics. *IEEE J Biomed Health Inform* 2022;26:1997–2007.
7 Fenech M, Strukelj N, BustonO. Ethical, social, and political challenges of artificial intelligence in health. The Wellcome Trust; 2018. 60.
8 Thamik H, Wu J. The impact of artificial intelligence on sustainable development in electronic markets. *Sustainability* 2022;14:3568.
9 Amann J, Blasimme A, Vayena E, *et al*. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak* 2020;20:310:310.:.
10 Lysaght T, Lim HY, Xafis V, *et al*. AI-assisted decision-making in healthcare: the application of an ethics framework for big data in health and research. *Asian Bioeth Rev* 2019;11:299–314.
11 Saraswat D, Bhattacharya P, Verma A, *et al*. Explainable AI for healthcare 5.0: opportunities and challenges. *IEEE Access* 2022;10:84486–517.
12 Nancy AA, Ravindran D, Raj Vincent PMD, *et al*. Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. *Electronics* 2022;11:2292.
13 Miotto R, Wang F, Wang S, *et al*. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinformatics* 2018;19:1236–46.
14 Rajkomar A, Dean J, Kohane I. Machine learning in medicine. Reply. *N Engl J Med* 2019;380:2589–90.
15 Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. *JAMA* 2018;320:2199–200.
16 Pawar U, O'Shea D, Rea S, *et al*. Incorporating explainable artificial intelligence (XAI) to aid the understanding of machine learning in the healthcare domain. AICS; 2020.
17 Shaban-Nejad A, Michalowski M, Buckeridge DL. Precision health and medicine. In: *Explainable AI in Healthcare and Medicine*. Cham: Springer, 2020.
18 Cutillo CM, Sharma KR, Foschini L, *et al*. Machine intelligence in healthcare-perspectives on trustworthiness, explainability, usability, and transparency. *NPJ Digit Med* 2020;3:47.
19 Laato S, Tiainen M, Najmul Islam AKM, *et al*. How to explain AI systems to end users: a systematic literature review and research agenda. Internet Res 2022;32:1–31.
20 Haque AB, Islam AKMN, Mikalef P. Explainable artificial intelligence (XAI) from a user perspective: a synthesis of prior literature and Problematizing avenues for future research. Technol Forecast Soc Change 2023;186:122120.
21 Shneiderman B. Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Trans Intell Syst* 2020;10:1–31.
22 Zanca F, BrusascoC, PesapaneF, *et al*. Regulatory aspects of the use of artificial intelligence medical software. *Semin Radiat Oncol* 2022.
23 Buolamwini J, Gebru T. Gender shades: Intersectional accuracy disparities in commercial gender classification. Conference on fairness, accountability and transparency; New York, NY, USA. PMLR, 2018
24 Mittelstadt BD, Allo P, Taddeo M, *et al*. The ethics of algorithms: mapping the debate. Big Data Soc 2016;3:205395171667967.
25 Hoadley KA, Yau C, Hinoue T, *et al*. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* 2018;173:291–304.