Editorial	*Open Access*

Urgent need for better quality control, standards and regulation for the Large Language Models used in healthcare domain

Brijesh Sathian[1], Edwin van Teijlingen[2], Israel Junior Borges do Nascimento[3], Russel Kabir[4], Indrajit Banerjee[5], Padam Simkhada[6], Hanadi Al Hamad[1]

Correspondence: Dr. Brijesh Sathian, Deputy Chair for Research, Geriatrics and long-term care department, Rumailah Hospital, Hamad Medical Corporation, Doha, Qatar.
Email: drsathian@gmail.com

Background

Globally, the increasing elderly population is associated with a rising prevalence of chronic diseases [1]. While advancements in medical care have extended life expectancy, there has not been a commensurate increase in healthy life years. This trend is escalating healthcare costs and compelling governments, insurers, individuals, families, regulators, and healthcare providers to seek innovative solutions and reform healthcare delivery systems worldwide [2]. Furthermore, in the context of the global health crisis, medical systems face the dual challenge of improving performance (delivering effective, high-quality care) while also transforming care delivery on a large scale by incorporating real-world, data-driven insights into patient treatment.

Artificial intelligence (AI) is fundamentally transforming the medical field, with the potential to enhance both physician and patient experiences. Although the transformative impact of AI in healthcare has been anticipated for years, only recently have technological advancements emerged that can capture some of the complexities of health, illness, and healthcare delivery [3, 4]. While large language models (LLMs) have demonstrated remarkable capabilities, the standards for clinical applications remain high. Efforts to evaluate these models' clinical knowledge often rely on automated assessments using narrow criteria. The recent prominence of LLMs in highly visible and interactive applications has generated interest in how emerging AI technologies might benefit medicine and health for patients, the public, physicians, health systems, and other stakeholders [5]. Critical areas of focus should include clinical care and outcomes, patient-centered care, healthcare quality, fairness in AI algorithms, medical education and clinician experience, and global solutions [6].

The potential of artificial intelligence to transform clinical research is substantial, with key advancements required in discovery science and medical applications. Research conducted by Liang et al. has demonstrated that large language models can provide valuable input on research papers [7]. While expert feedback remains essential for precise research, the rapid

increase in scholarly output has raised concerns regarding the efficacy of traditional scientific review processes. Obtaining high-quality peer reviews for prestigious journals is becoming increasingly challenging. Studies have revealed a strong correlation between LLM and human feedback in both retrospective and prospective evaluations, as well as positive user perceptions of LLM input's value [7]. Although human expert assessments should continue to be central to scientific processes, Large Language Model (LLM) contributions can support researchers, particularly when timely expert feedback is unavailable or during early manuscript preparation stages. The field of medical artificial intelligence (MAI) has evolved from traditional machine learning to deep learning, progressing towards unsupervised learning models [8]. Recent years have witnessed a shift from task-specific to generalized medical artificial intelligence (GMAI) models. These emerging AI models and algorithms need to be adapted for therapeutic applications across various contexts.

In the past five years, multiple research efforts have demonstrated that MAI models can match or surpass physician performance across various medical tasks and specialties [9-12]. Nevertheless, many of these models have been evaluated only through retrospective analyses using proxy outcomes, without considering real-world clinical environments.

Health information ecosystem data standards play a crucial role in facilitating software integration among healthcare entities for purposes such as data sharing, analysis, clinical research, and public health. However, the potential utilization of large language models to dynamically convert unstructured data into standardized formats for future use raises questions regarding the future of health data [13]. The regulation of clinical artificial intelligence (AI) presents unique challenges for global policymakers [14]. Current methodologies for ensuring AI technology safety and efficacy may be adequate for earlier AI iterations predating generative artificial intelligence (GAI). However, governing clinical GAI may necessitate the development of novel regulatory frameworks. As AI technology advances, researchers, academic institutions, funding bodies, and publishers should continue to examine its impact on scientific inquiry and revise their understanding, ethical guidelines, and regulations accordingly.

Acknowledgement

Competing interests

There is no conflict of interest for any author of this manuscript.

Authors' affiliations:

[1]Geriatrics and long term care department, Rumailah Hospital, Hamad Medical Corporation, Doha, Qatar

[2]Centre for Midwifery& Women's Health, Bournemouth University, Bournemouth, UK

[3]Division of Country Health Policies and Systems, World Health Organization Regional Office for Europe, Copenhagen, 2100, Denmark

[4]School of Allied Health, Anglia Ruskin University, Essex, UK

[5]Sir Seewoosagur Ramgoolam Medical College, Belle Rive, Mauritius

[6]School of Human and Health Sciences, University of Huddersfield, Huddersfield, UK

References

1.World Health Organization. Ageing and Health [Internet]. World Health Organization. 2022. [cited 2024 June 10]. Available from: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

2. World Health Organization. Working for health and growth: Investing in the health workforce. 2016. [cited 2024 June 10]. Available from: http://apps.who.int/iris/bitstream/10665/250047/1/9789241511308-eng.pdf

3. Lindberg DA. Medical informatics/computers in medicine. JAMA. 1986 Oct 17;256(15):2120-2. PMID: 3531559. https://doi.org/10.1001/jama.256.15.2120 PMid:3531559

4. Dorr DA, Adams L, Embí P. Harnessing the Promise of Artificial Intelligence Responsibly. JAMA. 2023 Apr 25;329(16):1347-1348. doi: 10.1001/jama.2023.2771. PMID: 36972068. https://doi.org/10.1001/jama.2023.2771 PMid:36972068

5. Haupt CE, Marks M. AI-Generated Medical Advice-GPT and Beyond. JAMA. 2023 Apr 25;329(16):1349-1350. doi: 10.1001/jama.2023.5321. PMID: 36972070. https://doi.org/10.1001/jama.2023.5321 PMid:36972070

6. Khera R, Butte AJ, Berkwits M, Hswen Y, Flanagin A, Park H, Curfman G, Bibbins-Domingo K. AI in Medicine-JAMA's Focus on Clinical Outcomes, Patient-Centered Care, Quality, and Equity. JAMA. 2023 Sep 5;330(9):818-820. doi: 10.1001/jama.2023.15481. PMID: 37566406. https://doi.org/10.1001/jama.2023.15481 PMid:37566406

7.Liang W, Zhang Y, Cao H, Wang B, Daisy Yi Ding, Yang X, et al. Can Large Language Models Provide Useful Feedback on Research Papers? A Large-Scale Empirical Analysis. NEJM AI. 2024 Jul 17; https://doi.org/10.1056/AIoa2400196

8. Ma W, Sheng B, Liu Y, Qian J, Liu X, Li J, et al. Evolution of Future Medical AI Models - From Task-Specific, Disease-Centric to Universal Health. NEJM AI. 2024 Jul 12; https://doi.org/10.1056/AIp2400289

9. Rajpurkar P, Lungren MP. The Current and Future State of AI Interpretation of Medical Images. N Engl J Med. 2023 May

25;388(21):1981-1990. doi: 10.1056/NEJMra2301725. PMID: 37224199. https://doi.org/10.1056/NEJMra2301725

PMid:37224199

10. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, Mahendiran T, Moraes G, Shamdas M, Kern C, Ledsam JR, Schmid MK, Balaskas K, Topol EJ, Bachmann LM, Keane PA, Denniston AK. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health. 2019 Oct;1(6):e271-e297. doi: 10.1016/S2589-7500(19)30123-2. Epub 2019 Sep 25. Erratum in: Lancet Digit Health. 2019 Nov;1(7):e334. doi: 10.1016/S2589-7500(19)30160-8. PMID: 33323251.

https://doi.org/10.1016/S2589-7500(19)30160-8

PMid:33345807

11. Rajpurkar P, Lungren MP. The Current and Future State of AI Interpretation of Medical Images. N Engl J Med. 2023 May 25;388(21):1981-1990. doi: 10.1056/NEJMra2301725. PMID: 37224199. https://doi.org/10.1056/NEJMra2301725

PMid:37224199

12. Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. Lancet Digit Health. 2024 May;6(5):e367-e373. doi: 10.1016/S2589-7500(24)00047-5. PMID: 38670745; PMCID: PMC11068159.

https://doi.org/10.1016/S2589-7500(24)00047-5

PMid:38670745

13. Brat GA, Mandel JC, Matthew B.A. McDermott. Do We Need Data Standards in the Era of Large Language Models? NEJM AI. 2024 Jul 19; https://doi.org/10.1056/AIe2400548

14. Blumenthal D, Patel B. The Regulation of Clinical Artificial Intelligence. NEJM AI. 2024 Jul 12;

https://doi.org/10.1056/AIpc2400545