Application and Analysis of AI Large Models in Health Science Popularization Q&A and Medical Image Interpretation

Kaiye Zhou

{zky2542@smail.seig.edu.cn}

Guangzhou University of Software, Guangzhou, Guangdong, China

Abstract. Current large language and multimodal foundation models are transforming medical question answering from search engine–based information retrieval to model-centered, dialogic consultations. While this shift enhances interactivity and generative capability, it also increases the risk of inaccurate or misleading responses. Therefore, this paper organizes the task lineage, method paradigms, and evaluation points along two lines: "Multimodal medical image analysis-Text-based health science popularization questions - Safety, ethics, and compliance governance". The paper is organized according to the "Subject domain-Task domain-Governance domain", prioritizing high-quality review articles, policy regulatory documents, and studies with comparable evaluation results to support the closed-loop argument from "Technology -Application-Risk-Governance", summarizing key risks such as hallucinations, overstepping boundaries, bias, and privacy, and forming a system-level mitigation closed-loop. At the same time, it clearly deploys the task list for future research: evidence-driven and verifiable generation, workflow integration, continuous evaluation and monitoring, and auditable governance.

Keywords: Large language model; Medical Q&A; Multimodal foundational model; Privacy and security governance.

1 Introduction

One of the fundamental needs in the medical information service chain is health science communication services and medical imaging interpretation: on the one hand, health knowledge elucidation and behavioral counseling to the general population; on the other hand, the semantic analysis and communicative interpretation of medical-imaging items in the service of diagnostic medicine. In the recent past, the use of large language models (LLM) in the medical field has gained momentum. This is not only in the field of medical education and medical information services but also clinical communication and research assistance. There are plenty of discussions and summary of reviews in this arena that offer clear guidelines to the development of problem decomposition and a design of a problem evaluation dimension [1, 2]. At the same time, AI applications aimed at the general population or at students have been known to have the potential of efficiency, but have a double-edged sword nature whereby, when abused as a

ICIAAI 2026, May 22-24, Singapore, Singapore

DOI 10.4108/eai.22-5-2026.2365084

base of diagnosis or treatment, can result in a delay of medical assistance or a wrong choice. Thus, solutions to both risk control and governance should be developed at the same time [3-5]. Since the medical imaging tasks will shift to the trend of a vision-language integrated report generation, image question-answering, and interpretive dialogue, therefore, when the visual capabilities are introduced, the cross-task generalization can be promoted. In the event that visual capabilities also can be trained, this facilitates a more substantial and improved quality of textual reportings. Nevertheless, it has negative aspects like bad semantic alignment, low interpretive consistency and propagation of errors [6].

The non-specialist health science communication Q&A covered herein is meant to assist the non-specialist users with clear and understandable explanations, practical given information and risk warnings- not as a substitute of a diagnosis or prescription. Medical image analysis involves both visual tasks and report generation, as well as image Q&A, which also is discussed with respect to its interactions with the system components, i. e., retrieval-enhanced retrieval, tool invocation, policy gating, audit monitoring, and compliance with privacy [5, 7].

The paper presents a two-track review framework as: medical image analysis-health science communication question-answer-security and ethical governance, presenting the technological development and priorities of multimodal and text applications evaluation. It provides a summary of the issue of governance in terms of bias, privacy, and regulation on an axis of risk in order to offer an auditable system-level mitigation loop with regards to which deployable systems could be referred to.

2 Route One: Application and Evolution of Medical Image Analysis and Health Science Q&A

2.1 Multimodal: Applications and Challenges of Large Models in Medical Image Analysis and Conversational Interpretation

The current progress of medical imaging AI has transformed the conventional feature engineering to deep learning applications and even vision-language/multimodal underlying models. In addition to classification, detection and segmentation, the tasks are also becoming less about generating reports, answering imaging questions and reasoning in the form of conversation with a greater focus on evidence and its clinical relevance to be incorporated into workflow [8]. To address clinical diagnostics, the evaluation metrics should include the visual indicators, clinical semantic consistency, and the level of safety/risk of an assessment. In the context of AI implemented in the benefit of the general population, warnings/alerts, explanations that recognize uncertainty should be included in deciphering publicly presented information. This prevents the risk of smoother explanations contributing to the risk of misinformation [6, 9, 10].

2.2 The Application and Evolution of Large Language Models in Health Science Q&A and Patient Education

The major objectives of text-based services are patient education and health promotion, consultation of symptoms and risk stratification, and interpretation of test results and reports by the lay population. Such services require the same care in terms of accuracy, comprehensibility

and actionability [1, 4, 11]. Such methodologies have developed across rule-based retrieval and discriminative learning to systematic ones such as RAG, tool invocation, and policy gating. One of the trends entails holding generation down on evidence of high authority and offering citations to be verified to minimize hallucinations and boundary transgression to source [5, 7]. The output that faces the general public must take the standard form (conclusion-cause-recommendation-red flags-information gaps), enhanced by explanations cues, rejection cues, and medical advice. The metrics of evaluation must include the harm severity grading, boundaries compliance grading and quality of refusal response [3, 12, 13].

3 Route Two: Safety, Ethics, and Compliance Governance

3.1 The Importance and Objectives of Governance

Keeping the cost of medical application errors is high, and governance is a systemic activity: there needs to be access control through the entire data lifecycle, retention limited, audits providing, and the goals of reducing severe misses, improving verifiability, privacy compliance, and auditable non-disclosure all require achievement.

3.2 Risk Spectrum

The main risks are the hallucinations, calibration risks, out of bounds and harmful risks, unverifiable interpretations risks, privacy and security risks and robustness and adversarial risks. Most of them include privacy and security threat as a decisive impediment to implementation in real practice, whereas diagnostic activities are typically characterized by overlapping hazards [5, 7, 10].

3.3 Deviation and Fairness: From “Performance Metrics” to “Population Impact”

The bias in AI can enhance the existing health disparities, and here the same system has different error patterns by age groups, gender groups, racial and geographical groups and even by patients with different comorbidities. Consequently, political correctness and justice are essential aspects in the analysis and implementation of immense medical models. This necessitates inclusion of subgroup performance reports, severity stratification of errors as well as continued auditing and development of mitigation strategies at three levels namely, data, model and interaction [14].

3.4 Regulatory Oversight and Boundaries of Responsibility: Why Need “Regulatory Discussions Tailored for LLMs”

Other theorists have come to believe that the regulation of LLM/generative AI in health care is needed: because the products of the activity of the LLM/generative AI do not exist as software programs or directly verifiable information, but in their generative processes, the generative AI has more serious opportunities and unpredictable limits than the traditional medical AI. The characteristics are peculiar and will require specific treatment, which can not be applied to existing certification and review approaches that are applied to currently medical AI [5]. Based on the credible standpoints as an educator, there is urgent need to define the roles of learning/ communication assistance and diagnostic/treatment recommendations to avoid a risk of using the generative contents as clinical fact [4].

3.5 System-Level Mitigation Strategy: From Single-Point Optimization to "Auditable Closed-Loop"

The following can be cited as the list of the system-level mitigation measures: Provide and generate citations against evidence/RAG, use structured templates to make boundary and risk disclosures, integrate evidence-based red team testing with uncertainty calibration, use policy gating and policy externalization of high-risk rules to facilitate auditable updates, provide human-machine collaborative review and continuous monitoring after deployment, form a closed-loop governance system.

4 Comprehensive Discussion: Advantages, Limitations, and Implementation Boundaries

The benefits of LLM are not only in the domain of natural language interpretation and interaction, knowledge services and medical education support, but also in the image interpretation + science communication model facilitated by multimodal fusion. Quick assessments based on special sets of questions and real world user scenarios are used to formulate the boundaries of its usability [15]. The limitations are mainly caused by hallucinations and overconfidence, untraceable evidence, and inconsistent explanation, out-of-domain generalization and distribution shift, bias and fairness, and privacy security and boundary regulatory responsibility. These include simultaneous development of evidence limits, evidence refusal reaction regulation, cross-centre validation, and information administration.

5 Future Development Trends and Research Agenda

The use of evidence chains, uncertainty statements, and risk grading as default output items may also become a standard part of communication in medical science and imaging interpretation in the future to limit the risks of hallucinations and boundary violations[5, 7]. The analysis of imagery and text-based question-answering will be enhanced in its role in pre- and post-visit processes and become the part of the service process model of consultation-examination-interpretation follow-up. This trend requires that not only must the models have multi-modal understanding, but also provide reasonable explanations, incorporate safety gates as well as be audit able[5, 9]. The current assessment will be replaced with a comprehensive lifecycle assessment that integrates the measures for their evaluation, which include naturally external validation + red team testing + post-deployment monitoring as the primary indicators of the quality, along with the final measures of the quality delivered by real users and the percentage of the security incidents as its indicators[5, 14]. Privacy compliance constrained, localized/private deployments, less retention of data and auditable governance will become more common. Problems in privacy and security offer easily understandable control techniques and paths of implementation of this direction of development [7]. In addition to Q&A and imaging, medical AI is coming together with predictive medicine, risk assessment, and clinical decision support systems, with increased requirements in AI data governance and trustworthiness. Multimodal foundational models will become a pillar of incorporating prediction-explanation-intervention recommendation of future AI systems [9, 16].

6 Conclusion

The given paper discusses the use opportunities of big language models. Approach One is a review of literature in the medical bioimage analysis and medical health science communication field and is structured or categorised into the types of tasks, methodology paradigms, and dimensions of evaluation in the order of multimodal-text. The process requires issues like a layered evaluation, expressions of safety to a public and limitations of evidence, and policy gating. One of these approaches is the focus on credible governance, which entails the compilation of risk-relevant undertakings that include privacy security, prejudice and equity, and regulatory control. Paired with this, it suggests auditable measures of system levels mitigation, stating that they are essential when large models, in this case, healthcare, enter into the healthcare domain. Lastly, it examines the future directions such as evidence-based generation, integration of multimodal workflows, real-world assessment and their application in privacy-friendly implementation. The ultimate is to come up with truly viable, trustworthy, and controllable medical science communication and image interpretation systems. trustworthy and controllable' medical science communication and imaging interpretation systems.

References

[1] Dave, T., Athaluri, S. A., Singh, S.: ChatGPT in medicine: An overview of its applications, advantages, limitations, future prospects, and ethical considerations. Frontiers in Artificial Intelligence. 6, Art. no. 1169595 (2023)
[2] Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., et al.: Large language models in medicine. Nature Medicine. 29(8), pp. 1930–1940 (2023)
[3] Boscardin, C. K., Gin, B., Golde, P. B., et al.: ChatGPT and generative artificial intelligence for medical education: Potential impact and opportunity. Academic Medicine. 99(1), pp. 22–27 (2024)
[4] Cooper, A., Rodman, A.: AI and medical education—A 21st-century Pandora's box. New England Journal of Medicine. 389(5), pp. 385–387 (2023)
[5] Meskó, B., Topol, E. J.: The imperative for regulatory oversight of large language models (or generative AI) in healthcare. npj Digital Medicine. 6(1), Art. no. 120 (2023)
[6] Koga, S., Du, W.: From text to image: Challenges in integrating vision into ChatGPT for medical image interpretation. Neural Regeneration Research. 20(2), pp. 487–488 (2025)
[7] Chen, Y., Esmaeilzadeh, P.: Generative AI in medical practice: In-depth exploration of privacy and security challenges. Journal of Medical Internet Research. 26, Art. no. e53008 (2024)
[8] Castiglioni, I., Rundo, L., Codari, M., et al.: AI applications to medical images: From machine learning to deep learning. Physica Medica. 83, pp. 9–24 (2021)
[9] Rajpurkar, P., Lungren, M. P.: The current and future state of AI interpretation of medical images. New England Journal of Medicine. 388(21), pp. 1981–1990 (2023)
[10] Ullah, E., Parwani, A., Baig, M. M., et al.: Challenges and barriers of using large language models such as ChatGPT for diagnostic medicine with a focus on digital pathology: A scoping review. Diagnostic Pathology. 19(1), Art. no. 43 (2024)
[11] Fattah, F. H., Salih, A. M., Salih, A. M., et al.: Comparative analysis of ChatGPT and Gemini (Bard) in medical inquiry: A scoping review. Frontiers in Digital Health. 7, Art. no. 1482712 (2025)
[12] Aydin, S., Karabacak, M., Vlachos, V., et al.: Large language models in patient education: A scoping review of applications in medicine. Frontiers in Medicine. 11, Art. no. 1477898 (2024)

[13] Friederichs, H., Friederichs, W. J., März, M.: ChatGPT in medical school: How successful is AI in progress testing? Medical Education Online. 28(1), Art. no. 2220920 (2023)
[14] Mittermaier, M., Raza, M. M., Kvedar, J. C.: Bias in AI-based models for medical applications: Challenges and mitigation strategies. npj Digital Medicine. 6(1), Art. no. 113 (2023)
[15] Gurbuz, S., Bahar, H., Yavuz, U., et al.: Comparative efficacy of ChatGPT and DeepSeek in addressing patient queries on gonarthrosis and total knee arthroplasty. Arthroplasty Today. 33, Art. no. 101730 (2025)
[16] Sharma, A., Lysenko, A., Jia, S., et al.: Advances in AI and machine learning for predictive medicine. Journal of Human Genetics. 69(10), pp. 487–497 (2024)