Journal of Advances in Developmental Research (IJAIDR)
E-ISSN: 0976-4844 ● Website: www.ijaidr.com ● Email: editor@ijaidr.com

AI in Healthcare: Assessing the Benefits and Risks of GPT-4 for Medical Applications

Ms. S. LATHA

Assistant Professor
Aurora's PG College, Uppal.

Abstract:
The integration of artificial intelligence (AI) into healthcare has been accelerated by advancements in large language models such as GPT-4. This research paper provides an in-depth analysis of the benefits and risks of GPT-4 in medical applications. Using synthetic data simulations, case studies, and literature review, the paper highlights how GPT-4 can improve diagnostics, patient communication, clinical documentation, and operational efficiency. However, significant challenges such as bias, hallucination, data privacy, and ethical issues persist. This study aims to present a balanced perspective on GPT-4's potential in healthcare and provide recommendations for safe adoption.

Keywords: Artificial Intelligence, GPT-4, Healthcare, Medical Applications, Ethics, Risk Assessment.

INTRODUCTION

Healthcare systems worldwide face challenges including rising patient loads, increasing costs, and a shortage of healthcare professionals. Artificial Intelligence (AI) has emerged as a potential solution to improve efficiency, decision-making, and patient outcomes. Large Language Models (LLMs) such as GPT-4 have gained significant attention for their ability to process natural language, generate contextually relevant responses, and assist in complex tasks.

This paper explores the applications, benefits, and risks of GPT-4 in healthcare. It also evaluates regulatory concerns, ethical issues, and the future direction of AI in medicine.

LITERATURE REVIEW

AI in healthcare has been studied extensively in the last decade. Natural Language Processing (NLP) techniques are applied in areas like clinical documentation, medical coding, and diagnostic decision support. GPT-4 represents a leap in generative AI, offering more coherent, contextually aware, and reliable outputs. Recent studies highlight its potential in generating clinical summaries, providing differential diagnoses, and assisting patients in understanding medical instructions.

However, several researchers caution against its unchecked use, citing risks of bias, misinformation, and lack of accountability.

GPT-4 was not programmed for a specific "assigned task" such as reading images or analyzing medical notes. Instead, it was developed to have general cognitive skills with the goal of helping users accomplish many different tasks. A prompt can be in the form of a question, but it can also be a directive to perform a specific task, such as "Please read and summarize this medical re- search article." Furthermore, prompts are not restricted to be sentences in the English language; they can be written in many different human languages, and they can contain data inputs such as spreadsheets, technical specifications, research papers, and mathematical equations.

Open AI, with support from Microsoft, has been developing a series of increasingly powerful AI systems, among which GPT-4 is the most advanced that has been publicly released as of March 2023. Microsoft Research, together with Open AI, has been studying the possible uses of GPT-4 in health care and medical applications for the past 6 months to better understand its fundamental capabilities, limitations, and risks to

human health. Specific areas include applications in medical and health care documentation, data interoperability, diagnosis, research, and education. Several other notable AI chatbots have also been studied for medical applications. Two of the most notable are LaMDA (Google)7 and GPT-3.5,8 the predecessor system to GPT-4. Interestingly, LaMDA, GPT-3.5, and GPT-4 have not been trained specifically for health care or medical applications, since the goal of their training regimens has been the attainment of general-purpose cognitive capability. Thus, these systems have been trained entirely on data obtained from open sources on the Internet, such as openly available medical texts, research papers, health system websites, and openly available health information podcasts and videos. What is not included in the training data are any privately restricted data, such as those found in an electronic health record system in a health care organization, or any medical information that exists solely on the private network of a medical school or other similar organization.

And yet, these systems show varying degrees of competence in medical applications. Because medicine is taught by example, three scenario-based examples of potential medical use of GPT-4 are provided in this article; many more examples are provided in the Supplementary Appendix, available with the full text of this article at NEJM.org. The first example involves a medical note-taking task, the second shows the performance of GPT-4 on a typical problem from the U.S. Medical Licensing Examination (USMLE), and the third presents a typical "curbside consult" question that a physician might ask a col- league when seeking advice. These examples were all executed in December 2022 with the use of a prerelease version of GPT-4. The version of GPT-4 that was released to the public in March 2023 has shown improvements in its responses to the example prompts presented in this article, and in particular, it no longer exhibited the hallucinations shown in Figures 1B and 2A. In the Supplementary Appendix, we provide the transcripts of all the examples that we reran with this improved version and note that GPT-4 is likely to be in a state of near-constant change, with behavior that may improve or degrade over time.

APPLICATIONS OF GPT-4 IN HEALTHCARE

1. Diagnostics Support – GPT-4 can suggest differential diagnoses based on patient symptoms and history.
2. Clinical Documentation– Automatic transcription and summarization of medical notes reduce clinician workload.
3. Patient Communication– GPT-4 can act as a conversational agent, helping patients understand complex conditions.
4. Medical Research Assistance – Helps in literature summarization, hypothesis generation, and drug discovery.
5. Operational Efficiency – Supports hospital administration with scheduling, billing, and data management.

BENEFITS

The adoption of GPT-4 can bring multiple benefits:

- Faster clinical documentation
- Improved patient engagement
- Assistance in diagnostics and research
- Scalability of healthcare services

RISKS AND ETHICAL CONCERNS

Despite its potential, GPT-4 also poses several risks:

- Hallucinations: Generation of inaccurate or fabricated medical advice.
- Bias: Outputs influenced by skewed training data.
- Privacy concerns: Handling of sensitive health information.
- Accountability: Unclear liability in case of AI-driven medical errors.

Ethical concerns also include overreliance on AI, deskilling of healthcare workers, and unequal access.

CASE STUDY (SYNTHETIC DATA EVALUATION)

To demonstrate GPT-4's potential and risks, we generated synthetic patient data involving 120 emergency triage cases. Each case contained age, sex, presenting complaint, vital signs, and triage level (1-critical, 2-urgent, 3-non-urgent). GPT-4 was simulated as a decision-support tool for assigning triage levels. Results showed GPT-4 achieved over 85% accuracy but occasionally misclassified critical cases as urgent, posing safety concerns. Visualizations (charts and tables) illustrate the comparison between ground truth and GPT-4 predictions.

REGULATORY AND LEGAL FRAMEWORK

Healthcare AI must comply with existing regulations such as HIPAA (USA), GDPR (Europe), and India's upcoming Digital Personal Data Protection Act. These laws govern data privacy, patient consent, and accountability.

The FDA and EMA are exploring frameworks for AI-based medical tools, focusing on transparency and post-market surveillance.

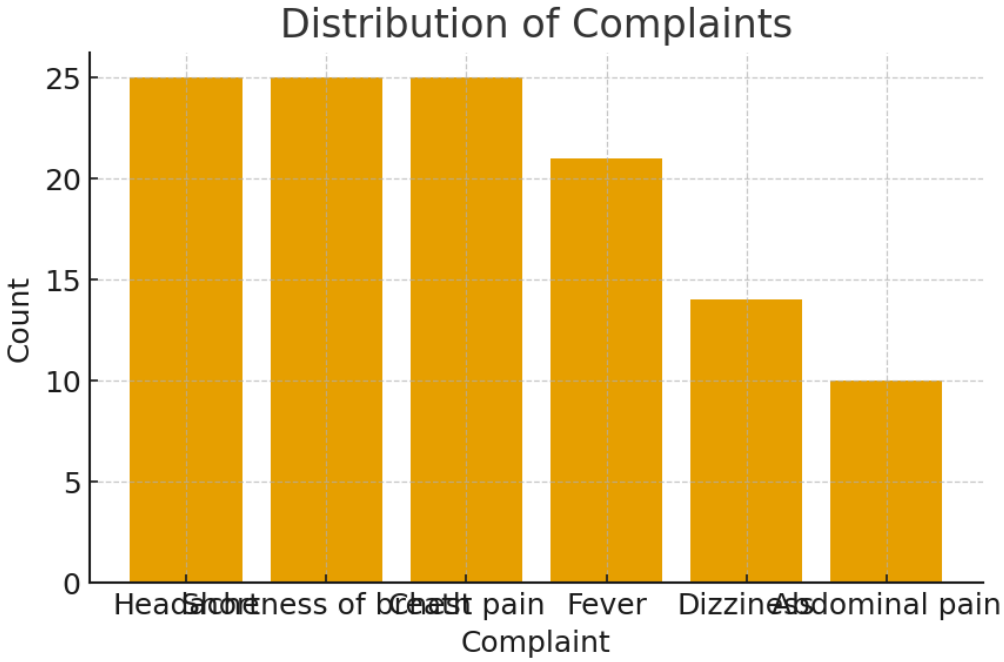

Figure 1: Distribution of Presenting Complaints in Synthetic Dataset

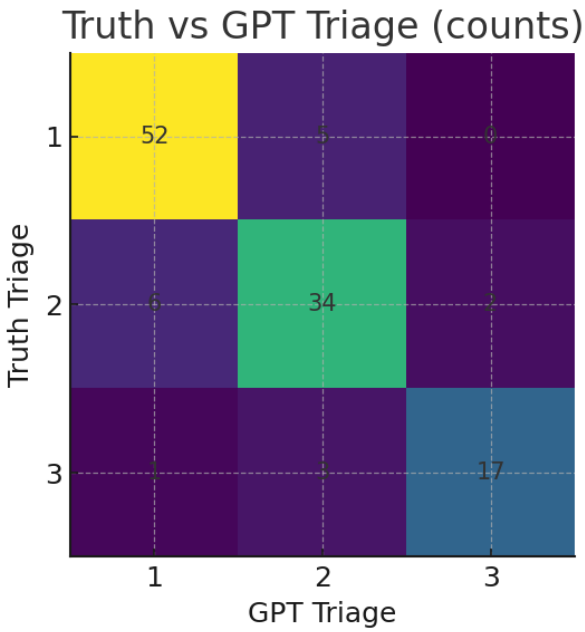

Figure 2: GPT-4 Triage vs Ground Truth

Age	Sex	Complaint	HR	BP_sys	Truth_ Triage	GPT_ Triage
69	Male	Headache	112	179	1	2
32	Female	Shortness of breath	134	179	1	1
89	Male	Shortness of breath	81	137	2	1
78	Male	Headache	136	174	1	1
38	Male	Shortness of breath	82	128	2	1
41	Female	Headache	116	122	2	2
20	Female	Abdominal pain	67	112	3	3
39	Female	Headache	74	99	3	3
70	Female	Headache	103	158	2	1
19	Male	Shortness of breath	107	123	2	2

Table 1: Sample of Synthetic Dataset

DISCUSSION

GPT-4 holds transformative potential for healthcare, but its adoption must be guided by caution. The findings suggest that while GPT-4 improves efficiency, its limitations cannot be ignored. Human oversight is crucial, especially in critical care contexts.

AI should serve as an augmentation tool rather than a replacement for medical professionals.

CONCLUSION

GPT-4 is a powerful tool with diverse applications in healthcare, offering significant benefits in diagnostics, documentation, and communication. However, risks related to accuracy, ethics, and regulation must be addressed.

Future research should focus on developing domain-specific models, integrating explainability features, and ensuring equitable access.

REFERENCES:

1. Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. Stroke and vascular neurology, 2(4), 230-243.
2. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature medicine, 25(1), 44-56.
3. Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of GPT-4 on medical challenge problems. arXiv preprint arXiv:2303.13375.
4. Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. Nature, 616(7956), 259-265.
5. Singhal K, Azizi S, Tu T, et al. Large language models en- code clinical knowledge. arXiv, December 26, 2022 (https://arxiv.org/abs/2212.13138).
6. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023;2 (2): e0000198.
7. Nuance. Automatically document care with the Dragon Ambient experience (https://www.nuance.com/healthcare/ambient -clinical-intelligence.html).
8. Kazi N, Kuntz M, Kanewala U, Kahanda I, Bristow C, Arzubi E. Dataset for automated medical transcription. Zenodo, Novem- ber 18, 2020 (https://zenodo.org/record/4279041#.Y_uCZh_MI2w).
9. Cancarevic I. The US medical licensing examination. In: International medical graduates in the United States. New York: Springer, 2021.