

Access this article online
Quick Response Code:

Website: www.jehp.net
DOI: 10.4103/jehp.jehp_688_23

A large language model in solving primary healthcare issues: A potential implication for remote healthcare and medical education

Himel Mondal, Rajesh De¹, Shaikat Mondal², Ayesha Juhi

Abstract:

BACKGROUND AND AIM: Access to quality health care is essential, particularly in remote areas where the availability of healthcare professionals may be limited. The advancement of artificial intelligence (AI) and natural language processing (NLP) has led to the development of large language models (LLMs) that exhibit capabilities in understanding and generating human-like text. This study aimed to evaluate the performance of a LLM, ChatGPT, in addressing primary healthcare issues.

MATERIALS AND METHODS: This study was conducted in May 2023 with ChatGPT May 12 version. A total of 30 multiple-choice questions (MCQs) related to primary health care were selected to test the proficiency of ChatGPT. These MCQs covered various topics commonly encountered in primary healthcare practice. ChatGPT answered the questions in two segments—one is choosing the single best answer of MCQ and another is supporting text for the answer. The answers to MCQs were compared with the predefined answer keys. The justifications of the answers were checked by two primary healthcare professionals on a 5-point Likert-type scale. The data were presented as number and percentage.

RESULTS: Among the 30 questions, ChatGPT provided correct responses for 28 yielding an accuracy of 93.33%. The mean score for explanation in supporting the answer was 4.58 ± 0.85 . There was an inter-item correlation of 0.896, and the average measure intraclass correlation coefficient (ICC) was 0.94 (95% confidence interval 0.88–0.97) indicating a high level of interobserver agreement.

CONCLUSION: LLMs, such as ChatGPT, show promising potential in addressing primary healthcare issues. The high accuracy rate achieved by ChatGPT in answering primary healthcare-related MCQs underscores the value of these models as resources for patients and healthcare providers in remote healthcare settings. This can also help in self-directed learning by medical students.

Keywords:

Artificial intelligence, ChatGPT, large language model, multiple-choice questions, primary healthcare, remote health care, telemedicine

Introduction

Access to quality health care is crucial for individuals worldwide, but it can be particularly challenging in remote areas where healthcare resources are limited. The emergence of artificial intelligence (AI) and natural language processing (NLP) has opened new avenues for improving

healthcare delivery in underserved regions. Large language models (LLMs), such as ChatGPT, powered by advanced AI algorithms, have shown remarkable capabilities in understanding and generating human-like text.^[1,2]

In remote areas, the scarcity of healthcare professionals poses a significant barrier to quality care.^[3] Utilizing AI-powered

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Mondal H, De R, Mondal S, Juhi A. A large language model in solving primary healthcare issues: A potential implication for remote healthcare and medical education. J Edu Health Promot 2024;13:362.

Department of Physiology,
All India Institute of
Medical Sciences,
Deoghar, Jharkhand,
India, ¹Department of
Community Medicine,
Malda Medical College
and Hospital, Malda,
West Bengal, India,
²Department of Physiology,
Raiganj Government
Medical College and
Hospital, Raiganj,
West Bengal, India

Address for correspondence:

Dr. Ayesha Juhi,
Department of Physiology,
All India Institute of
Medical Sciences,
Deoghar - 814 152,
Jharkhand, India.
E-mail: ayesha.physiology
@aiimsdeoghar.edu.in

Received: 18-05-2023
Accepted: 17-07-2023
Published: 28-09-2024

language models like ChatGPT may help bridge this gap by providing accurate and reliable healthcare information to patients and even assisting healthcare providers in making informed decisions. By leveraging the vast knowledge and language understanding abilities of ChatGPT, primary healthcare concerns can be addressed remotely, potentially improving healthcare outcomes and reducing the burden on limited healthcare resources. Healthcare providers in remote locations can utilize ChatGPT as a supportive tool, allowing them to enhance their knowledge base and provide more comprehensive care to their patients. These implications highlight the potential of LLMs in transforming the landscape of remote health care and improving healthcare access for underserved populations.^[4]

In health care, accurate information is vital for proper diagnosis, treatment recommendations, and patient management. Healthcare professionals and patients alike rely on accurate and up-to-date knowledge to make informed decisions about medical conditions, medications, and interventions.^[5] The accuracy of a LLM is of utmost importance when considering its use in health care. Healthcare decisions are critical and can have a direct impact on patient outcomes. Therefore, relying on an LLM with high accuracy ensures that the information provided is reliable and trustworthy, reducing the risk of errors or misinformation.^[6]

In this context, this pilot study aimed to explore the effectiveness of ChatGPT, a LLM, in addressing primary healthcare issues. By evaluating its accuracy in answering a set of primary healthcare-related multiple-choice questions (MCQs), we can assess its potential role in providing healthcare information and support to individual case scenario. The findings of this study can contribute to know the potential of integration of AI-powered language models into telemedicine platforms, leading to enhanced healthcare access and improved health outcomes for underserved populations.

Materials and Methods

Study type and setting

This was a cross-sectional study conducted in May 2023 on the World Wide Web with the help of a personal computer connected to the Internet by a personal broadband connection. No institutional resources were used in this study.

Study participants and sampling

A set of 30 primary healthcare-related MCQs was compiled for this study. These questions were compiled by two primary healthcare experts. The MCQs were also checked by another primary care physician and

were pretested. These MCQs covered a range of topics commonly encountered in primary healthcare practice.

The LLM used in this study was ChatGPT developed by OpenAI.^[7] ChatGPT is trained on a diverse range of Internet text and has demonstrated proficiency in understanding and generating human-like responses. There is a free version of the application that can be accessed for research purposes (ChatGPT May 12 Version).

The 30 MCQs were presented to ChatGPT, and their responses were saved for further analysis. The questions had a common string—"Answer the question by choosing the single best answer from the provided options with a justification of the answer." ChatGPT provided the answer in two parts—the correct option among the provided four options and text in support of the answer (i.e., justification). The justification or explanation text was stored for further analysis. Figure 1 presents the study flow, and Figure 2 presents a screenshot captured during a conversation with ChatGPT.

Assessment of accuracy

The accuracy of the answers was compared to the known correct answers for each question. The accuracy of ChatGPT's responses was calculated by determining the percentage of questions it answered correctly. The justification text was analyzed by two primary care experts who scored the responses on a 5-point Likert-type scale where 1 = completely inaccurate,

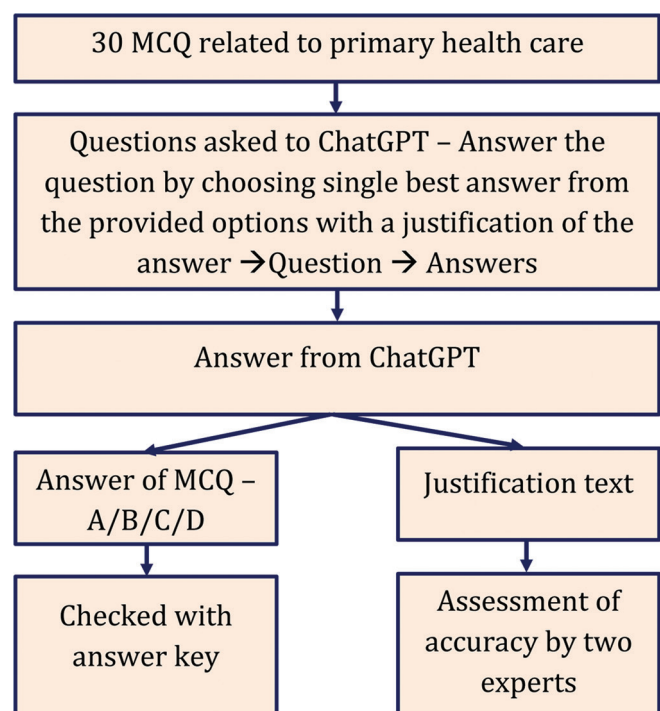


Figure 1: Study process in brief

2 = significant inaccuracies, 3 = mix of accurate and inaccurate information, 4 = indicates mostly accurate, and 5 = highly accurate.

Statistical analysis

The data were presented in number and percentage. The central tendency was measured by mean and standard deviation. The agreement between the score of the two observers was tested by the intraclass correlation coefficient (ICC). Statistical tests were conducted from available online tools suggested by Mondal *et al.*^[8] We considered a *P* value <0.05 to be statistically significant.

Ethical considerations

This study was conducted in compliance with ethical guidelines. No real patient's data were used in this study. The dataset used for evaluation was anonymized and did not contain any personally identifiable information.

Results

Among the 30 case scenarios presented to the ChatGPT, the correct response was received for 28 cases that indicate the accuracy of the model to be 93.33% for solving primary healthcare issue. Among the domains, one answer was wrong in promotive primary care, one answer was wrong in preventive primary care, and there was no error in curative primary care.

The first observer rated the response 4.6 ± 0.93 and the second observer rated the response 4.57 ± 0.82 with an average of two as 4.58 ± 0.85 [Table 1]. This indicates that, on average, respondents rated the explanation to be accurate.

There was a strong inter-item correlation of 0.896, indicating that the items or statements assessing the

quality of the explanation were consistently related to each other. The Bland–Altman plot is shown in Figure 3. The average measure ICC was 0.94, with a 95% confidence interval of 0.88 to 0.97. This indicates a high level of interobserver agreement among the individuals assessing the explanations.

Discussion

The study aimed to assess the accuracy of a LLM—ChatGPT in solving primary healthcare issues, specifically through answering MCQs related to primary health care. The results showed that the LLM demonstrated a high level of accuracy, correctly answering 28 of 30 MCQs. Hence, ChatGPT may serve as a decision support tool, assisting remote healthcare providers in accessing quick and reliable information. This may help in improving diagnostic accuracy and treatment decisions. Previous studies also explored its potential to solve problems in pathology, microbiology, and pharmacology with an average of 80% accuracy.^[9–12] Additionally, the LLM's accuracy makes it a reliable source of healthcare information for a wider audience, promoting health literacy and empowering individuals to make informed decisions.^[13] Even patients, who do not have immediate access to healthcare services, may get help from the language model.

Table 1: Central tendency of scores awarded by two raters and their correlation

	Rater 1	Rater 2	Average	<i>r, P</i>
Mean	4.6	4.57	4.58	0.896,
Standard deviation	0.93	0.82	0.85	<0.0001
Range	1-5	2-5	1.5-5	
First quartile–third quartile	5-5	4-5	4.5-5	

Answer the question by choosing the single best answer from the provided options with a justification of the answer.

Question: A 27-year-old man comes in with a complaints of fever, dry cough, headache, and body ache for two days. On examination, the fever is 101°F, heart rate 114 bpm, blood pressure 132/88 mmHg, mild redness but no pus exudates in the throat. Systemic examination was not remarkable. What kind of management is best at this point?

Answer options:

- a) Oral Azithromycin
- b) Oral Amoxicillin
- c) Oral Paracetamol
- d) Oral Levocetirizine

The most appropriate management option at this point would be c) Oral Paracetamol.

Justification:

Based on the patient's symptoms (fever, dry cough, headache, and body ache), along with the absence of pus exudates in the throat and unremarkable systemic examination, it is likely that the patient is presenting with a viral respiratory infection rather than a bacterial infection. The presence of mild redness in the throat suggests a possible viral pharyngitis.

Figure 2: An example of a conversation with ChatGPT to solve the primary care health issue

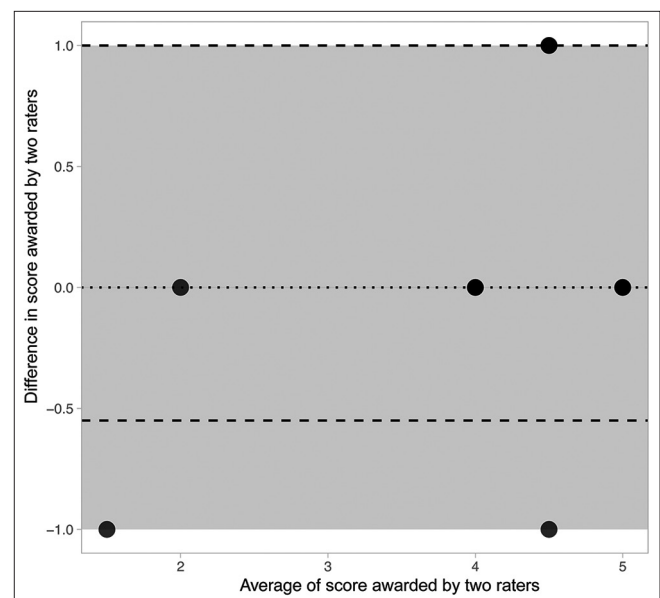


Figure 3: Bland–Altman plot for the scores awarded by two raters

Developing countries often face limitations in resources and infrastructure, making it difficult to access up-to-date and reliable healthcare information. LLM can serve as a valuable tool, providing a wealth of medical knowledge and evidence-based information to healthcare professionals and individuals in remote areas. Furthermore, LLM can act as a decision support system, helping healthcare providers in resource-constrained settings where access to specialized expertise may be limited.^[14] It can provide recommendations, treatment guidelines, and diagnostic insights, thereby assisting healthcare professionals in making accurate and timely decisions.^[15] In our study, among the questions we tested, no wrong answer was for curative primary care.

In addition to helping in healthcare delivery, LLM can contribute to training and education in developing countries by offering a vast repository of medical knowledge, clinical guidelines, and case studies. In medical education, self-directed learning is an important mode of learning where students direct themselves to learning a topic. The ChatGPT can help medical students by providing justification for the answer we used in this study.^[16] This can aid in the professional development of healthcare workers and improve their skills.

However, it is important to acknowledge the limitations of LLMs. Despite their higher level of accuracy, LLMs are still reliant on the quality and comprehensiveness of the data they are trained on. The success of LLMs in healthcare applications also depends on the appropriate design of the system, including considerations for privacy, ethics, and the integration of human expertise.^[17] The language model may not provide up-to-date information as it depends on the training data that are usually not real time. In addition, AI is yet far away from human empathy and emotional intelligence that is an important aspect of medical teaching.^[18,19] Further research is needed to address these challenges and optimize the use of LLMs in healthcare settings.

Conclusion

This study contributes to the growing body of literature on the use of LLMs in healthcare. The findings highlight the high accuracy of an LLM in solving primary healthcare issues. These results suggest that LLMs have the potential to support remote health care and serve as a valuable tool for healthcare professionals. However, further research and development are required to fully leverage the capabilities of LLMs and address the associated challenges to ensure their safe and effective integration into healthcare practice.

Acknowledgement

The authors thank all the primary care physicians who reviewed the questions for checking their accuracy and content validity. The authors also thank the participants in the pretesting of the questions. The first author would like to thank Sarika Mondal and Ahana Aarshi for sacrificing their family time to allow data collection, analysis, and preparation of this manuscript. Some portions of the article were edited for language and grammar by ChatGPT free research preview (May 24 version).

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Arora A, Arora A. The promise of large language models in health care. *Lancet* 2023;401:641. doi: 10.1016/S0140-6736 (23) 00216-7.
2. Tustumi F, Andreollo NA, Aguilar-Nascimento JE. Future of the language models in healthcare: The role of ChatGPT. *Arq Bras Cir Dig* 2023;36:e1727. doi: 10.1590/0102-672020230002e1727.
3. Sharma DC. India still struggles with rural doctor shortages. *Lancet* 2015;386:2381-2. doi: 10.1016/S0140-6736 (15) 01231-3.
4. Sallam M. ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)* 2023;11:887. doi: 10.3390/healthcare11060887.
5. Alotaibi YK, Federico F. The impact of health information technology on patient safety. *Saudi Med J* 2017;38:1173-80. doi: 10.15537/smj.2017.12.20631.
6. Moore AR, Hudson C, Amey F, Chumbler N. Trusting sources of information on quality of physician care. *Inquiry* 2020;57:46958020952912. doi: 10.1177/0046958020952912.
7. Lubowitz JH. ChatGPT, an artificial intelligence chatbot, is impacting medical literature. *Arthroscopy* 2023;39:1121-2. doi: 10.1016/j.arthro.2023.01.015.
8. Mondal S, Mondal H, Panda R. How to conduct inferential statistics online (Part 2): A brief hands-on guide for biomedical researchers. *Indian J Vasc Endovasc Surg* 2022;9:63-9. doi: 10.4103/ijves.ijves_130_21.
9. Sinha RK, Deb Roy A, Kumar N, Mondal H. Applicability of ChatGPT in assisting to solve higher order problems in pathology. *Cureus* 2023;15:e35237.
10. Das D, Kumar N, Longjam LA, Sinha R, Deb Roy A, Mondal H, *et al.* Assessing the capability of ChatGPT in answering first- and second-order knowledge questions on microbiology as per competency-based medical education curriculum. *Cureus* 2023;15:e36034.
11. Juhi A, Pipil N, Santra S, Mondal S, Behera JK, Mondal H. The capability of ChatGPT in predicting and explaining common drug-drug interactions. *Cureus* 2023;15:e36272.
12. Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, *et al.* Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin Mol Hepatol* 2023. doi: 10.3350/cmh.2023.0089.

13. Khan RA, Jawaaid M, Khan AR, Sajjad M. ChatGPT-Reshaping medical education and clinical management. *Pak J Med Sci* 2023;39:605-7. doi: 10.12669/pjms. 39.2.7653.
14. Shanbehzadeh M, Nopour R, Kazemi-Arpanahi H. Developing an artificial neural network for detecting COVID-19 disease. *J Educ Health Promot* 2022;11:2.
15. Chow JCL, Sanders L, Li K. Impact of ChatGPT on medical chatbots as a disruptive technology. *Front Artif Intell* 2023;6:1166014. doi: 10.3389/frai.2023.1166014.
16. Lee H. The rise of ChatGPT: Exploring its potential in medical education. *Anat Sci Educ* 2023. doi: 10.1002/ase.2270.
17. Gupta A, Singla T, Chennatt JJ, David LE, Ahmed SS, Rajput D. Artificial intelligence: A new tool in surgeon's hand. *J Educ Health Promot* 2022;11:93.
18. Sheikhbardsiri H, Sheikhasadi H, Mahani SA, Mohamadi MMD. Emotional intelligence and learning strategies of postgraduate students at Kerman University of Medical Sciences in the southeast of Iran. *J Educ Health Promot* 2020;9:66. doi: 10.4103/jehp.jehp_544_19.
19. Sarani A, Mousavi SH, Salahi S, Hasani F, Abdar ZE, Sheikhbardsiri H. Emotional intelligence and self-efficacy among deputy's administrative staff of Kerman University of Medical Sciences. *J Educ Health Promot* 2020;9:105. doi: 10.4103/jehp.jehp_482_19.