Evaluating ChatGPT's Educational Suitability for Patient Primary Care Queries

Monica Khadka, BS, BA | Riya Rupareliya, BS | Deepali Khadka, MD | Ajay Bisht, MD

Abstract

Introduction: ChatGPT, a large language model created by OpenAI, has emerged as a new source of online medical information. This study aimed to evaluate the appropriateness, readability, and educational value of ChatGPT's responses to frequent patient internet queries regarding 10 common primary care diagnoses.

Methods: The responses generated by ChatGPT regarding the 10 most frequently encountered primary care diagnoses were assessed for appropriateness and readability by two primary care physicians. Responses were judged based on educational value in four categories: basic knowledge, diagnosis, treatment, and prevention. We used a 5-point Likert scale based on accuracy, comprehensiveness, and clarity to determine appropriateness. ChatGPT responses that received ratings of 4-5 in all three criteria were considered appropriate. Conversely, if the outputs received ratings of 1-3 in any category, they were deemed inappropriate. We performed readability assessments using the Flesch Reading Ease (FRE) and Flesch-Kincaid Reading Grade Level (FKGL) formulas to determine if the responses were at the recommended average American's seventh to eighth grade reading level.

Results: Most (92.5%) responses were deemed appropriate unanimously by both reviewers. ChatGPT provided more appropriate responses regarding basic knowledge compared to diagnosis, treatment, and prevention. The ChatGPT responses demonstrated a college graduate reading level, as indicated by the mean FRE score of 25.64 and the median FKGL score of 12.61.

Conclusion: Our comprehensive analysis found that ChatGPT's responses were appropriate most of the time. These findings suggest that ChatGPT has potential to be a supplementary educational tool for patients seeking health information online.

Introduction

ChatGPT is a growing artificial intelligence (AI) tool gaining popularity due to its application in various fields, including medicine. As a large language model (LLM), ChatGPT is trained on vast data sets of text to generate human-like responses to user input through pattern recognition and probabilistic modeling.[1] Several studies have explored the applicability and accuracy of ChatGPT responses across medical topics. Recent research has evaluated ChatGPT's ability to generate patient-facing educational materials in emergency medicine, surgical settings, and other clinical domains.[2-4]

One study found that ChatGPT outperformed Google in answering common questions related to breast implant-associated anaplastic large cell lymphoma.[5] While many papers investigate ChatGPT's usefulness across specialized fields, few have explored its ability to address questions related to primary care diagnoses.[5, 6] Established studies have examined ChatGPT's medical knowledge base, with one study showing that ChatGPT was at or nearly at the passing threshold for all three United States Medical Licensing Exams (USMLE).[7] These studies suggest ChatGPT may serve as a medical educational tool.[7] This study aims to determine ChatGPT's suitability as a source of medical information for patients, defined by the appropriateness (accuracy, comprehensiveness, and clarity) and readability of its responses, with the broader goal of improving patient health literacy.

Methods

This cross-sectional study evaluated the appropriateness and readability of responses generated by ChatGPT regarding 10 commonly-encountered primary care diagnoses: pneumonia, osteoarthritis, congestive heart failure, sepsis, atrial fibrillation, chronic obstructive pulmonary disease (COPD), depression, hypertension, diabetes, and gastroesophageal reflux disease (GERD). Diagnoses were selected based on informal literature review[8] and the clinical judgment of two board-certified primary care physicians.

Standardized patient-focused questions were developed and inputted into ChatGPT (GPT-4, OpenAI). Responses were judged based on educational value in four categories: basic knowledge ("What is [diagnosis]?"), diagnosis ("What are the signs and symptoms of [diagnosis]?"), treatment ("How do I treat [diagnosis]?"), and prevention ("How can I prevent [diagnosis]?"). Basic knowledge was defined as a general explanation of the condition. Each diagnosis was queried with four questions, yielding 40 total responses. No initial instructions were given regarding reading level. However, when follow-up requests were given for simplified language, appropriate responses were generated.

We used a 5-point Likert scale assessing accuracy, comprehensiveness, and clarity to determine appropriateness, with higher scores indicated better performance. Responses that received ratings of 4-5 in all three criteria were considered appropriate. If a response received a rating of 1-3 in any category from either reviewer, it was deemed inappropriate. Ratings were based on clinical expertise and comparison to evidence-based guidelines (UpToDate, Centers for Disease Control and Prevention).

Readability was assessed using the Flesch Reading Ease (FRE) and Flesch-Kincaid Reading Grade Level (FKGL) formulas to evaluate alignment with the recommended sixth to seventh grade reading level. FRE scores range from 0 (very difficult) to 100 (very easy), and FKGL scores range from 0 (learning to read books) to 18 (academic paper). Descriptive statistics were used to summarize appropriateness and readability outcomes. Ratings were analyzed in aggregate across all diagnoses; individual scores by diagnosis were not retained in a stratified format.

Results

Of the 40 responses analyzed, 35 (87.5%) were rated as appropriate by both reviewers. Responses addressing diagnosis received the highest appropriateness rate (100%), followed by prevention (90%), treatment (80%), and basic knowledge (80%). Treatment had two low-rated responses—one for comprehensiveness and one for accuracy. Basic knowledge had two low-rated responses—one for comprehensiveness and one for clarity. Prevention had one low-rated response for comprehensiveness. No low-rated responses were found in the diagnosis category.

Readability analysis revealed that ChatGPT's responses often exceeded the average US reading level. The mean

FRE score was 25.64 and the median FKGL score was 12.61, corresponding to a college graduate reading level. While the responses were typically clear and well-structured, they frequently included advanced medical terminology, such as "angiotensin receptor blockers" and "inflammatory cytokines," without adequate simplification into layman's terms, unless prompted to do so. A sample response for pneumonia is presented in Table 1.

Conclusions

ChatGPT is an AI tool gaining attention in the medical field for its potential to improve medical education and access to information. While most existing studies focus on ChatGPT's usefulness from the clinician's standpoint, its relevance from the patient's viewpoint remains unexplored. This study among the first to examine the patient education quality of responses related to 10 common primary care diagnoses. Results showed 92.5% of all responses deemed appropriate unanimously by two primary care physicians, suggesting that ChatGPT may be a supplementary educational tool for patients seeking health information online. ChatGPT's speed and ease of use may make it impactful in-patient education.[9] By improving access to health care information, ChatGPT has the potential to increase health literacy while addressing health care disparities.

Several limitations should be noted. The small sample size limits the generalizability, and future studies with larger datasets are needed. The use of only two physician reviewers may introduce subjectivity, although both reviewers independently rated each response, and only those deemed appropriate by both were classified as such.

Readability was also a concern. The mean FRE score was 25.64 and the median FKGL was 12.61, indicating a college-level reading complexity. This poses challenges for patients with lower literacy levels, as the average reading level in the United States is estimated to be between the seventh and eighth grade.[10] FRE and FKGL are proxies for textual complexity and were used to estimate accessibility for the public. ChatGPT's outputs may vary based on question phrasing, and the potential for "hallucinations"—factually incorrect responses that sound realistic—is a known limitation of LLMs.[11] Clinician oversight may help mitigate these risks. Further, our Likert scale may not fully capture acceptability, and the prompts used may reflect a provider lens. Future studies should incorporate patient-informed questions to better reflect typical patient concerns. Our study contributes to current literature by focusing on patient-facing information for common ambulatory primary care conditions and suggests that AI may be a promising tool for improving patient medical literacy.

Tables and Figures

<p align="center">**Table 1: A Sample ChatGPT Response for Pneumonia***</p>

Category	Details
Basic knowledge	Pneumonia is an inflammatory condition of the lung affecting primarily the small air sacs known as alveoli. It is commonly caused by bacterial, viral, fungal, or other infections. Pneumonia can also result from inhaling chemical irritants or aspirating (inhaling) food or other objects into the lungs.
Signs and symptoms/diagnosis	• Cough: pneumonia often starts with a productive cough that may produce phlegm. • Shortness of breath: difficulty breathing or rapid breathing may occur. • Chest pain: pain or discomfort may be felt in the chest, especially when breathing or coughing. • Fatigue: general weakness or tiredness is common. • Confusion: in older adults, pneumonia can lead to confusion or changes in mental awareness. Diagnosing pneumonia typically involves a combination of: • Medical history: Your doctor will inquire about your symptoms, overall health, and any recent exposure to factors that could cause pneumonia. • Physical examination: This may include listening to your lungs with a stethoscope to detect abnormal sounds. • Chest X-ray: This imaging test helps to confirm the presence and extent of lung involvement. • Blood tests: These may be done to identify the specific cause of the infection, such as bacteria or viruses.
Treatment	The treatment of pneumonia depends on the underlying cause: • Bacterial pneumonia: Antibiotics are often prescribed to combat bacterial infections. • Viral pneumonia: Antiviral medications may be used for certain viral infections, but supportive care such as rest and hydration is also crucial. • Fungal pneumonia: Antifungal medications are used for fungal infections. • Symptom relief: Medications to alleviate symptoms like fever and cough may be recommended. It's crucial to complete the prescribed course of antibiotics or antiviral medications, even if symptoms improve before the medication is finished.
Prevention	• Vaccination: Vaccines are available for some of the most common causes of pneumonia, including influenza (flu) and pneumococcus. • Hand hygiene: Wash hands regularly to prevent the spread of germs. • Respiratory hygiene: Cover your mouth and nose with a tissue or your elbow when coughing or sneezing • Avoid smoking: Smoking damages the lungs and increases the risk of respiratory infections. • Healthy lifestyle: Maintain a healthy diet, exercise regularly, and get adequate rest to support overall immune function.
Note	If you suspect you have pneumonia or are experiencing severe respiratory symptoms, seek medical attention promptly. Early diagnosis and appropriate treatment are essential for a successful recovery.

* Content generated using ChatGPT (OpenAI, 2024).[12]

Corresponding Author

Riya Rupareliya, BS

Florida State University College of Medicine, Tallahassee, FL

Rr22u@fsu.edu

Author Affiliations

Monica Khadka, BS, BA - University of South Florida Health Morsani College of Medicine, Tampa, FL

Riya Rupareliya, BS - Florida State University College of Medicine, Tallahassee, FL

Deepali Khadka, MD - University of Florida Health, Gainesville, FL

References

1. Open AI. ChatGPT (version GPT-4) [Large language model]. 2024. Accessed July 15, 2025. https://chat.openai.com

2. Halaseh FF, Yang JS, Danza CN, Halaseh R, Spiegelman L. ChatGPT's role in improving education among patients seeking emergency medical treatment. *West J Emerg Med*. 2024;25(5):845-855. doi:10.5811/WESTJEM.18650

3. Abdelmalek G, Uppal H, Garcia D, Farshchian J, Emami A, McGinniss A. Leveraging ChatGPT to produce patient education materials for common hand conditions. *J Hand Surg Glob Online*. 2024;7(1):37-40. doi:10.1016/j.jhsg.2024.10.002

4. Wei Q, Yao Z, Cui Y, Wei B, Jin Z, Xu X. Evaluation of ChatGPT-generated medical responses: A systematic review and meta-analysis. *J Biomed Inform*. 2024;151:104620. doi:10.1016/j.jbi.2024.104620

5. Liu HY, Alessandri Bonetti M, De Lorenzi F, Gimbel ML, Nguyen VT, Egro FM. Consulting the digital doctor: google versus ChatGPT as sources of information on breast implant-associated anaplastic large cell lymphoma and breast implant illness. *Aesthetic Plast Surg*. 2024;48(4):590-607. doi:10.1007/s00266-023-03713-4

6. Endo Y, Sasaki K, Moazzam Z, et al. Quality of ChatGPT responses to questions related to liver transplantation. *J Gastrointest Surg*. 2023;27(8):1716-1719. doi:10.1007/s11605-023-05714-9

7. 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. doi:10.1371/journal.pdig.0000198

8. National Center for Health Statistics. National Ambulatory Medical Care Survey: 2019 National Summary Tables. Centers for Disease Control and Prevention; 2021. Accessed July 12, 2025. https://www.cdc.gov/nchs/data/ahcd/namcs_summary/2019-namcs-web-tables-508.pdf

9. Clark M, Bailey S. Chatbots in health care: connecting patients to information: emerging health technologies. Canadian Agency for Drugs and Technologies in Health. Published January 2024. Accessed January 9, 2025. https://www.ncbi.nlm.nih.gov/books/NBK602381/

10. Doak CC, Doak LG, Friedell GH, Meade CD. Improving comprehension for cancer patients with low literacy skills: strategies for clinicians. *CA Cancer J Clin*. 1998;48(3):151-162. doi:10.3322/canjclin.48.3.151

11. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *J Med Syst*. 2023;47(1):33. doi:10.1007/s10916-023-01925-4

12. Open AI. GPT-4 Technical Report. 2023. Accessed July 12, 2025. https://openai.com/research/gpt-4