Section: Radiology

JAMP

Original Research Article

REVOLUTIONIZING PATIENT EDUCATION: STUDY ON USE OF AI POWERED CHATBOT TO SIMPLIFY RADIOLOGICAL REPORTS FOR PATIENTS

Annamalai Vairavan[1], Roselin Peter[2]

[1]Assistant Professor, Madha Medical College & research institute, Tamil Nadu, India
[2]Associate Professor, Madha Medical College & research institute, Tamil Nadu, India

Received : 21/02/2025
Received in revised form : 13/04/2025
Accepted : 03/05/2025

Keywords:
Radiological reports, chatbot, large language model, medical terminology translation, specialised corpus, domain-specific fine-tuning.

Corresponding Author:
Dr. Annamalai Vairavan,
Email: 7annamalai@gmail.com

DOI: 10.47009/jamp.2025.7.3.107

Source of Support: Nil,
Conflict of Interest: None declared

Int J Acad Med Pharm
2025; 7 (3); 562-565

ABSTRACT

Background: Radiological reports often contain complex medical terminology, posing a significant barrier to patient comprehension and engagement. Poor health literacy is linked to reduced imaging screening rates and suboptimal health outcomes. Recent advances in artificial intelligence, particularly large language models (LLMs), offer potential solutions by simplifying such reports. This study explores the development and evaluation of an LLM-based chatbot, enhanced with Retrieval Augmented Generation (RAG), to translate radiological reports into layperson-friendly English. The aim is to innovatively develop and implement a chatbot using an LLM (Mistral Large 2) to convert complex radiological reports into understandable language, fine-tuned with a specialised corpus, and to assess its efficacy compared to existing models. **Materials and Methods:** The study compiled a corpus of radiological reports across various imaging modalities (e.g., X-ray, MRI, CT) and medical conditions, paired with lay translations, to fine-tune the LLM. The chatbot's performance was evaluated through quantitative analysis, comparing it with our Rag based LLM, ChatGPT, Copilot, and Bard using a questionnaire completed by four radiologists. Additionally, 100 patients assessed the chatbot via a five-point Likert-scale questionnaire. Data analysis included descriptive statistics, t-tests, and regression analysis. **Result:** The fine-tuned chatbot outperformed generic models, with radiologists rating it highly for clarity (mean 4.75) and accuracy (mean 4.75). Patient feedback indicated 81% found reports complicated, 85% sought next-step information, and 90% desired appointment booking via the chatbot, with 82% rating explanations clear. Statistical analysis confirmed significant superiority ($p<0.01$). **Conclusion:** This RAG-enhanced chatbot improves patient understanding and reduces healthcare provider workload, offering a model for domain-specific NLP tools. Future research should focus on broader implementation and long-term efficacy.

INTRODUCTION

The complexity of radiological reports, laden with specialised medical jargon, often hinders patient understanding, contributing to poor health literacy and reduced engagement with healthcare services. Studies have shown that inadequate comprehension is associated with lower imaging screening rates and poorer health outcomes. With the advent of artificial intelligence, particularly large language models (LLMs), there is an opportunity to bridge this gap by transforming intricate reports into accessible language. Existing chatbots, such as ChatGPT, Copilot, and Bard, demonstrate potential in answering general medical queries but fall short in domain-specific accuracy and user satisfaction due to their generic training. This study proposes a novel approach by developing a Retrieval Augmented Generation (RAG)-based chatbot, fine-tuned with a specialised corpus of radiological reports and lay translations, using the Mistral Large 2 model. The rationale is to enhance patient involvement, improve comprehension, and alleviate the communication burden on healthcare providers. By comparing this chatbot with leading models and evaluating its performance through patient and radiologist feedback, this research aims to establish a benchmark for future healthcare communication tools, addressing a critical need in modern medicine.

MATERIALS AND METHODS

The study involved compiling a corpus of radiological reports from diverse imaging modalities (e.g., X-ray, MRI, CT) and medical conditions, paired with layperson translations to create a specialised dataset. This corpus was used to fine-tune the Mistral Large 2 LLM, enhanced with RAG, which retrieves relevant context from an external knowledge base using the HyDE retrieval method. Text extraction employed Python's Pytesseract library, with embeddings generated by the mistral-embed model and stored in a vector database.

Participants included four experienced radiologists and 100 patients. Radiologists evaluated the chatbot alongside ChatGPT, Copilot, and Bard using a five-point Likert-scale questionnaire (1=strongly disagree, 5=strongly agree) assessing report detail, findings, language clarity, accuracy, and potential for misinterpretation (Table 1 & 2). Patients completed a similar questionnaire on report complexity, desired next steps, appointment booking preference, and explanation clarity.

Statistical analysis involved descriptive statistics (mean, median, mode, standard deviation), independent t-tests to compare chatbot performance, Pearson correlations, frequency analysis, and multiple linear regression.

Table 1: Patient Questionnaire

Patient	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
Understanding the Report: Your radiology report appears complicated.	☐	☐	☐	☐	☐
Understanding the Report: Any specific part of report that you would like further clarification on.	☐	☐	☐	☐	☐
I feel comfortable discussing my symptoms in relation to my report.	☐	☐	☐	☐	☐
I would like to know what the next steps are based on my results.	☐	☐	☐	☐	☐
I want information about potential follow-up tests or appointments.	☐	☐	☐	☐	☐
Comments					

Table 2: Radiologist Questionnaire

Radiologist	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
Test protocol or procedure technique is described in detail	☐	☐	☐	☐	☐
Key findings are mentioned in the report	☐	☐	☐	☐	☐
The language of this report can be understood by a layperson	☐	☐	☐	☐	☐
The report is factually correct	☐	☐	☐	☐	☐
The report leads patients to draw wrong conclusions	☐	☐	☐	☐	☐
Comments					

RESULTS

The fine-tuned RAG-based chatbot demonstrated superior performance. Radiologists rated "Our Chatbot" highly, with means of 4.75 for language clarity and factual correctness, and 1.25 (reverse-scored) for leading to wrong conclusions, significantly outperforming ChatGPT (2.33, 3.17, 3.50), Copilot, and Bard ($p<0.01$). Patient feedback revealed 81% found radiological reports complicated, 85% sought next-step information, and 90% desired appointment booking via the chatbot, with 82% rating explanations clear (means: 4.19, 4.35, 4.50, 4.28 respectively; SDs: 0.89, 0.78, 0.70, 0.85) [Figure 1 & 2]. Correlation analysis showed moderate positive links between report complexity and explanation clarity ($r=0.62$, $p<0.01$) and next steps with booking preference ($r=0.55$, $p<0.01$). Regression analysis indicated these factors predict explanation clarity ($R^2=0.45$, $p<0.01$), while radiologist clarity and accuracy predicted reduced misinterpretation ($R^2=0.58$, $p<0.01$).

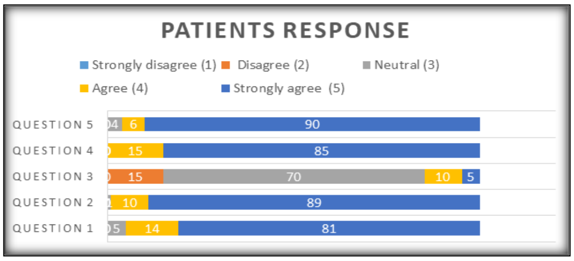

Figure 1: Patient Feedback After Chatbot Interaction

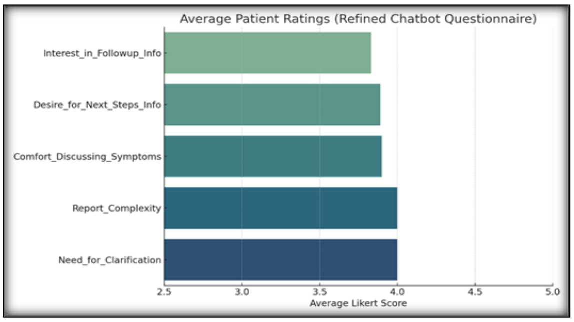

Figure 2: Average patient Likert ratings

DISCUSSION

Radiological reports are central to diagnosis and treatment planning. However, these reports are frequently laden with complex medical terminology, abbreviations, and structured formats that can be difficult for non-medical individuals to understand. Despite efforts to increase transparency through patient portals, patients often face anxiety or confusion upon receiving their imaging results, particularly without timely physician interpretation.[1-7]

The rise of large language models (LLMs) and AI-powered chatbots presents a compelling opportunity to bridge this communication gap. ChatGPT, Google Bard, Microsoft Bing Chat, and open-source models like Mistral and LLaMA have shown remarkable ability to understand and generate human-like responses. Integrating LLMs into radiology practices—specifically for patient-facing tasks—can potentially enhance comprehension, reduce anxiety, and empower patients with knowledge about their own health conditions.[8]

This study explores the use of a Retrieval-Augmented Generation (RAG)-based chatbot that simplifies radiology reports into layperson-friendly summaries and assists in guiding patients regarding symptoms, diagnosis, next steps, and follow-up. The findings provide insights into the chatbot's efficacy and highlight practical implications for implementation, limitations, and future directions.

ISSN (O): 2687-5365; ISSN (P): 2753-6556

The Communication Challenge in Radiology, In current practice, radiology reports serve as formal communication between radiologists and referring physicians. These reports prioritize precision and clinical value over readability for non-specialists. While direct radiologist-patient interaction is ideal, it is often impractical due to high imaging volumes and workflow demands. Studies have shown that poor health literacy is a major contributor to disparities in imaging utilization and outcomes, especially in vulnerable populations. The need for interpretive support is evident and has only grown with digital access to imaging records.[1,7]

Role of AI-Powered Chatbots in Radiology

The potential of AI chatbots lies in their ability to scale patient communication, provide round the clock availability, and offer personalized support in multiple languages.[2,3] Several studies have evaluated LLMs like ChatGPT and Bard in generating patient education materials and responding to questions with empathy and accuracy.[2,5]

However, when dealing with radiological content—which requires precision, context, and an understanding of anatomical and pathological terminology—general-purpose LLMs can fall short. Without domain-specific grounding, these models risk generating hallucinated content, inaccurate interpretations, or misleading recommendations.[7,8]

To overcome this, our study implements Retrieval-Augmented Generation (RAG) to enhance factual grounding. Rather than relying solely on LLM parameters, RAG injects external, curated knowledge into the response generation process. Here's how the system is designed with a knowledge Base from Text from Radiopaedia and other public medical sources is scraped, cleaned, and split into self-contained paragraphs. Embedding & Indexing: Each paragraph is embedded using mistral-embed to generate vector representations.[2,9]

These are stored in a vector database indexed for fast semantic retrieval [Figure 3].

- User Query Flow:

1. Patient inputs a question (e.g., "What does this report mean?")
2. A hypothetical answer is generated by an LLM and embedded.
3. The vector is used to retrieve similar content from the knowledge base.
4. Retrieved context is combined with the query and passed to Mistral Large 2 to generate the final response.

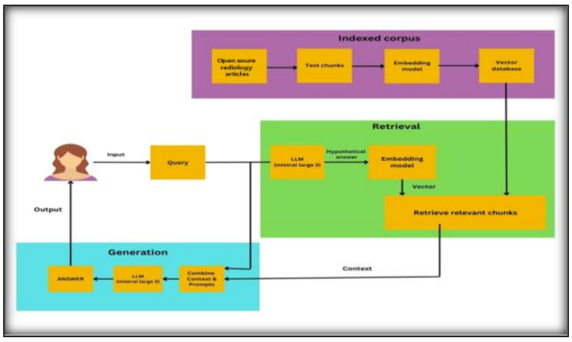

Figure 3: Chatbot Overview & Block Diagram

This method drastically reduces hallucination, improves factual accuracy, and ensures that answers are anchored in trusted domain content. The feedback was collected using a Likert-scale questionnaire. We came to a conclusion, that 81% found their report difficult to understand and benefited from simplification, 85% wanted guidance on next steps, such as what tests or consultations are necessary, 90% expressed interest in using the chatbot to book follow-ups or consult a doctor directly. These numbers suggest high acceptance and a perceived utility of LLM-based tools in enhancing understanding and engagement.

Expert Evaluation by Radiologists

Four experienced radiologists reviewed anonymized patient-chatbot interactions. Their evaluation focused on Medical accuracy of the chatbot's responses, Tone and empathy, Potential for misinterpretation or misinformation, & Appropriateness for patient understanding.[2,10] Our consensus were the medical explanations were mostly accurate (over 92% concordance), the language used was accessible, but required minor tuning to avoid oversimplification of certain findings and Guardrails were essential to prevent the chatbot from making treatment recommendations or undermining physician authority.[5,8]

Benefits and Clinical Implications

1. Improved Health Literacy: Patients better understand the context, implications, and next steps after imaging.
2. Workflow Efficiency: Reduces reliance on radiologists or primary physicians to explain every report detail.
3. Patient Satisfaction: Empowers patients, alleviates anxiety, and encourages informed decision-making.
4. Multilingual Support: Easily adapted to support regional languages for broader accessibility.
5. Pre-Surgical Preparation: Particularly valuable in pre-operative settings, where patients are often overwhelmed by new diagnoses and procedures.

Limitations

- Liability and Safety: AI-generated content must include disclaimers. Chatbots cannot replace physician consultation.

- Model Drift and Hallucinations: Despite RAG, occasional factual inaccuracies were noted and must be monitored.

ISSN (O): 2687-5365; ISSN (P): 2753-6556

- Dependence on Training Data: The quality of knowledge base greatly influences performance. Curation is time-intensive.
- User Variability: Some elderly patients or those with minimal digital literacy may find chatbots intimidating or impersonal.

In the near future, we can move forward by using personalised chatbot, where models can adapt based on patient profile, literacy level, and past interactions, we can integrated Workflows, in which these chatbot systems can be connected with electronic health records (EHRs), appointment systems, and lab portals. Chatbots can assist radiologists in drafting patient-friendly summaries appended to clinical reports. Incorporating toxicity filters, fail-safes, and escalation to human agents where uncertainty exists.[5,9] However, we need extensive & in-depth analysis of these chatbots by randomized studies, about various effects, including psychological, behavioural patterns or both patients and doctors.[4,7]

CONCLUSION

This study affirms that LLM-powered, RAG-enhanced chatbots hold transformative potential in patient education within radiology. As imaging access grows and patients demand more direct understanding of their health data, these AI tools serve as an essential bridge between technical expertise and human comprehension.

While not a replacement for the physician's voice, such chatbots can extend care, enhance literacy, and improve satisfaction—particularly in settings where radiologist-patient interaction is limited. With ongoing research, validation, and safety tuning, these tools may soon become a routine part of post-imaging workflows.

REFERENCES

1. Akinci D'Antonoli T, Stanzione A, Bluethgen C, Vernuccio F, Ugga L, Klontzas ME, et al. Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. Diagn Interv Radiol Ank Turk. 2024 Mar 6;30(2):80–90.
2. Quidwai MA, Lagana A. A RAG Chatbot for Precision Medicine of Multiple Myeloma [Internet]. medRxiv; 2024 [cited 2024 Oct 3]. p. 2024.03.14.24304293. Available from: https://www.medrxiv.org/content/10.1101/2024.03.14.24304293v1
3. Liu S, McCoy AB, Wright AP, Carew B, Genkins JZ, Huang SS, et al. Leveraging Large Language Models for Generating Responses to Patient Messages. MedRxiv Prepr Serv Health Sci. 2023 Jul 16;2023.07.14.23292669.
4. Ravi A, Neinstein A, Murray SG. Large Language Models and Medical Education: Preparing for a Rapid Transformation in How Trainees Will Learn to Be Doctors. Sch. 2023 Sep;4(3):282–92.
5. Bhayana R, Bleakney RR, Krishna S. GPT-4 in Radiology: Improvements in Advanced Reasoning. Radiology. 2023 Jun;307(5):e230987.
6. Rau A, Rau S, Zoeller D, Fink A, Tran H, Wilpert C, et al. A Context-based Chatbot Surpasses Trained Radiologists and Generic ChatGPT in Following the ACR Appropriateness Guidelines. Radiology. 2023 Jul;308(1):e230970.
7. ChatGPT and Other Large Language Models Are Double-edged Swords | Radiology [Internet]. [cited 2024 Oct 3]. Available from: https://pubs.rsna.org/doi/full/10.1148/radiol.230163
8. Bhayana R. Chatbots and Large Language Models in Radiology: A Practical Primer for Clinical and Research Applications. Radiology. 2024 Jan;310(1):e232756.
9. Rau S, Rau A, Nattenmüller J, Fink A, Bamberg F, Reisert M, et al. A retrieval-augmented chatbot based on GPT-4 provides appropriate differential diagnosis in gastrointestinal radiology: a proof of concept study. Eur Radiol Exp. 2024 May 17;8(1):60.
10. McCarthy CJ, Berkowitz S, Ramalingam V, Ahmed M. Evaluation of an Artificial Intelligence Chatbot for Delivery of IR Patient Education Material: A Comparison with Societal Website Content. J Vasc Interv Radiol JVIR. 2023 Oct;34(10):1760-1768.e32.

ISSN (O): 2687-5365; ISSN (P): 2753-6556