

## REVIEWER'S REPORT

Manuscript No.: IJAR-52123

Date: 06.06.2025

**Title: QUESTION-ANSWER SYSTEM ON MEDICAL DOMAIN WITH LLMS USING VARIOUS FINE – TUNING METHODS**

### Recommendation:

Accept as it is .....

✓ **Accept after minor revision.....**

Accept after major revision .....

Do not accept (*Reasons below*) .....

Rating	Excel.	Good	Fair	Poor
Originality		✓		
Techn. Quality		✓		
Clarity			✓	
Significance	✓			

Reviewer Name: Ms. S. Lavanya , AP/IT

Date: 06.06.2025

### Reviewer's Comment for Publication.

*(To be published with the manuscript in the journal)*

*The reviewer is requested to provide a brief comment (3-4 lines) highlighting the significance, strengths, or key insights of the manuscript. This comment will be Displayed in the journal publication alongside with the reviewers name.*

This manuscript presents a practical and timely approach to deploying large language models (LLMs) in the medical domain using efficient fine-tuning techniques such as QLoRA and PEFT. The integration of reinforcement learning with human feedback and multilingual capabilities enhances its applicability in real-world healthcare settings. The work demonstrates strong potential for improving clinical decision support systems

## REVIEWER'S REPORT

### *Detailed Reviewer's Report*

#### **“Question-Answer System on Medical Domain with LLMs Using Various Fine-Tuning Methods”**

### **1. General Assessment**

The manuscript presents a comprehensive approach to building a **voice-based question-answering (QA) system in the medical domain**, leveraging various **fine-tuning methods for large language models (LLMs)**, including **QLoRA, LoRA, PEFT, RLHF, and Chain-of-Thought prompting**. It proposes a low-resource, multilingual system designed for practical deployment using open-source LLMs and lightweight hardware, which is highly relevant for healthcare applications in low-resource settings.

### **2. Strengths**

- **Timeliness and Practicality:**  
The topic is extremely relevant due to the rising interest in LLMs and their applicability in domain-specific tasks such as medical diagnostics and QA systems.
- **Comprehensive Pipeline:**  
The proposed three-phase architecture (Speech-to-Text, Finetuning, Text-to-Speech) is clearly defined and well thought-out, supporting multilingual input and output.
- **Use of Efficient Techniques:**  
The paper demonstrates strong awareness of hardware constraints and employs **parameter-efficient fine-tuning methods** (e.g., QLoRA, PEFT) to train models effectively on modest computational setups.
- **Experimental Validation:**  
The authors evaluate various LLMs and techniques, including CoT prompting and ensemble learning, using real medical datasets (MedMCQA and USMLE). The results are informative and demonstrate the benefit of finetuning.

### **3. Areas for Minor Revision**

1. **Language and Grammar:**
  - There are several grammatical and typographical errors throughout the document (e.g., “Back Traslation” instead of “Back Translation”).
  - Some sentences are overly complex or awkwardly structured. A thorough proofreading is required.
2. **Formatting and Consistency:**
  - Variable formatting is inconsistent in figures and equations.
  - Citations (e.g., [30], [24]) should be consistently linked to the reference list and used inline where necessary.

# International Journal of Advanced Research

Publisher's Name: Jana Publication and Research LLP

*www.journalijar.com*

---

## REVIEWER'S REPORT

### 3. Clarity in Experimental Section:

- While the evaluation tables are useful, clearer explanations of the test setup, scoring method, and metrics used (especially in Table 4.x) would help the reader better understand the significance of results.

### 4. Dataset Usage Clarification:

- The data preparation process, including how questions were split and cleaned, is mentioned but could benefit from more transparency.

### 5. Enhance the Conclusion:

- The conclusion could be more impactful by better summarizing the performance insights and explicitly stating potential clinical implications or deployment strategies.

## 4. Recommendation

**Recommendation:** Accept after minor revision

## 5. Overall Ratings

Criteria	Rating
Originality	Good
Technical Quality	Good
Clarity	Fair
Significance	Excellent

## 6. Final Comment for Editor

This manuscript offers a timely and relevant contribution to the application of LLMs in medical question answering. Its focus on low-resource finetuning and multilingual support enhances its real-world value. With some editorial improvements, this paper will be a strong addition to the literature on applied AI in healthcare.