

REVIEWER'S REPORT

Manuscript No.: **IJAR-58423**

Title: QUESTION-ANSWER SYSTEM ON MEDICAL DOMAIN WITH LLMS USING VARIOUS FINE-TUNING & RAG WITH MCP 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's ID: JPR- **098**

Detailed Reviewer's Report

Decision: Accept after Major Revision

Summary:

The manuscript proposes a **medical domain question-answering system** based on **Large Language Models (LLMs)** using **QLoRA, PEFT, RLHF, Chain-of-Thought (CoT), ensemble prompting, and a Corrective Retrieval-Augmented Generation (CRAG) framework** to improve clinical reasoning and reduce hallucinations. The system is evaluated on the **MedMCQA and USMLE (MedQA)** datasets.

Strengths:

- Addresses a timely and important topic in **medical LLMs and clinical question answering**.
- Integrates several modern techniques including **QLoRA, PEFT, RLHF, CoT prompting, and CRAG** into a unified framework.
- Uses publicly available benchmark datasets (MedMCQA and USMLE/MedQA).
- Experimental evaluation includes zero-shot, CoT, ensemble prompting, and CRAG-based retrieval enhancement.

International Journal of Advanced Research

Publisher's Name: Jana Publication and Research LLP

www.journalijar.com

REVIEWER'S REPORT

- The proposed CRAG validation mechanism for reducing hallucinations has practical relevance for healthcare AI.

Major Issues:

- The manuscript combines many existing techniques (QLoRA, RLHF, PEFT, CoT, PPO, CRAG) but the **novel methodological contribution is not clearly distinguished** from prior work.
- Experimental evaluation is **limited to only 100 test questions**, which is insufficient to support the reported performance claims.
- No statistical significance analysis, confidence intervals, or multiple-run evaluation is provided.
- Comparisons are limited; the study does not benchmark against **recent medical LLMs** (e.g., Med-PaLM, BioMistral, Llama-3-based medical models).
- Several mathematical formulations (e.g., PPO loss equations) are incomplete or incorrectly formatted, reducing technical clarity.
- The reported CRAG improvements are promising, but the evaluation lacks sufficient details regarding retrieval settings, reproducibility, and computational overhead.
- Numerous references are **preprints (arXiv) or future-dated (2025–2026)** and should be carefully verified before publication.

Minor Issues:

- Several grammatical, typographical, and formatting errors are present throughout the manuscript.
- Some figures are screenshots rather than publication-quality illustrations.
- Equation formatting and notation should be standardized.
- Section numbering and figure references require consistency.
- The conclusion overstates the effectiveness of the proposed framework relative to the reported experimental evidence.

Final Comment:

The manuscript addresses an important application of LLMs in medical question answering and integrates several state-of-the-art fine-tuning and retrieval techniques into a unified framework. However, the work currently

International Journal of Advanced Research

Publisher's Name: Jana Publication and Research LLP

www.journalijar.com

REVIEWER'S REPORT

lacks sufficient methodological novelty, rigorous experimental validation, reproducibility details, and comprehensive benchmarking against recent medical LLMs. Addressing these issues would significantly strengthen the scientific contribution.