

---

Peer-Review Report

# Peer Review of “The Performance of DeepSeek R1 and Gemini 3 in Complex Medical Scenarios: Comparative Study”

---

Jacqueline Guan-Ting You

Mass General Brigham, Boston, MA, United States

---

**Related Articles:**

Preprint (medRxiv):: <https://www.medrxiv.org/content/10.1101/2025.04.29.25326666v1>

Authors' Response to Peer-Review Reports: <https://med.jmirx.org/2026/1/e96220>

Published Article: <https://med.jmirx.org/2026/1/e76822>

*JMIRx Med* 2026;7:e96225; doi: [10.2196/96225](https://doi.org/10.2196/96225)

---

**Keywords:** large reasoning model; LRM; large language model; LLM; accuracy; medical scenario; DeepSeek R1; Gemini 3

*This is a peer-review report for “The Performance of DeepSeek R1 and Gemini 3 in Complex Medical Scenarios: Comparative Study.”*

## Round 1 Review

---

### General Comments

This paper [1] seeks to evaluate the accuracy of DeepSeek R1 in correctly identifying the primary medical diagnosis in the medical scenarios dataset portion of Massive Multi-task Language Understanding Pro (MMLU-Pro) using an open-ended format. Some clarifications on the methods and results (especially around the roles of subject matter experts vs core team members in the publication), would be helpful in understanding how these results were derived.

### Specific Comments

#### Minor Comments

1. Introduction: consider citing Deepseek AI's Deepseek R1 paper [2].
2. Methods: please clarify who your subject matter experts were (eg, physicians, researchers) in terms of rank, specialty, and role and how they were used to grade answers (eg, selected based on specialty, 2 reviewer process, etc).

3. Methods: please indicate when the analyses were run.
4. Results: who determines whether references are related or unrelated?
5. Results and Discussion: it is unclear to me from reading the discussion portion of the paper as to whether we have any sense of whether DeepSeek R1 has correct reasoning for questions with correct diagnoses (eg, it may get the right diagnosis but may have incorrect reasoning). Similarly, did you determine the “correct answer” based on string matching (for example, if the answer was “septic arthritis” and the DeepSeek output stated “septic shock,” would this be incorrect)?
6. Discussion: consider acknowledging the sample size of questions as a limitation.

## Round 2 Review

---

### General Comments

The paper has been revised to address the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Large Language Model (TRIPOD-LLM) guidelines. Overall, it appears most concerns from both reviewers have been addressed.

---

**Conflicts of Interest**

None declared.

---

**References**

1. Bajwa M, Hoyt R, Knight D, Haider M. The performance of DeepSeek R1 and Gemini 3 in complex medical scenarios: comparative study. *JMIRx Med*. 2026;7:e76822. [doi: [10.2196/76822](https://doi.org/10.2196/76822)]
2. Guo D, Yang D, Zhang H, et al. DeepSeek-R1: incentivizing reasoning capability in LLMs via reinforcement learning. Preprint posted online on Jan 22, 2025. [doi: [10.48550/arXiv.2501.12948](https://doi.org/10.48550/arXiv.2501.12948)]

---

**Abbreviations**

**MMLU-Pro:** Massive Multitask Language Understanding Pro

**TRIPOD-LLM:** Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Large Language Model

---

*Edited by Amy Schwartz; This is a non-peer-reviewed article; submitted 26.Mar.2026; final revised version received 26.Mar.2026; accepted 26.Mar.2026; published 27.Apr.2026*

*Please cite as:*

*You JGT*

*Peer Review of “The Performance of DeepSeek R1 and Gemini 3 in Complex Medical Scenarios: Comparative Study”*

*JMIRx Med 2026;7:e96225*

*URL: <https://med.jmirx.org/2026/1/e96225>*

*doi: [10.2196/96225](https://doi.org/10.2196/96225)*

© Jacqueline Guan-Ting You. Originally published in JMIRx Med (<https://med.jmirx.org>), 27.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIRx Med, is properly cited. The complete bibliographic information, a link to the original publication on <https://med.jmirx.org/>, as well as this copyright and license information must be included.