3RD PLACE FOR HIGHEST SCORE DURING THE ABSTRACT REVISION PROCESS FOR ORIGINAL RESEARCH (4 WAY TIE!!!)

ORIGINAL RESEARCH

03. Assessing the Precision of AI-Generated Medical Answers: An Evaluation of LLaMA-3 Powered Meta AI

Areeba Khan[1], Muhammad Talha Khan[2]

[1]Liaquat National Hospital and Medical College, National Stadium Road, Karachi, 74800, Sindh, Pakistan

[2]Dow University of Health Sciences, Baba-e-Urdu Road, Karachi, 74200, Sindh, Pakistan

https://www.youtube.com/watch?v=hJIcIJ1w8oM&list=PLhqNq3xJClbafO0Y5bvBcgMmXpgzJxd44&index=5&t=1526s

Background & Objective: Meta AI is being used frequently by medical students to clear the queries or to solve the questions, due to its easy availability. The purpose of this study is to assess the correctness of the Meta AI-generated answers to medical questions, and the reproducibility of the results.

Method: The study employs an Evaluation Research Design aimed to assess the quality and effectiveness of Meta AI. A total of 240 MCQs were included in the questionnaire, 30 MCQs from each subject. Out of these, 108 were case-based (Category: A), whereas 132 were fact-based (Category: B). Meta AI was re-queried with the previously failed questions 14 days later. Results were analyzed manually and accuracies were evaluated using IBM SPSS version 27.

Results: In initial analysis, Meta AI correctly answered 187 out of 240 questions (average accuracy = 77.9%). The most accurately responded category was Category-A with an average accuracy of 82.4%. The accuracy of Category-B was noted to be 74.2%. In re-scored analysis, Meta AI reproduced correct answers for only 12 out of 53 previously failed questions leading to an average reproducibility of 22.6%. The accuracy of Category A was 47.3% and that of Category B was 8.8%.

Conclusion: The integration of AI in the field of medicine is advancing rapidly, and models like Meta AI represent significant strides in making medical information more accessible and accurate. Despite these promising results, there are notable limitations, such as the scope of questions, the subjects covered, and potential selection biases.

Table 1. Overview of Total and Categorical Validation from Initial Analysis

Subject	Correct Answers	Wrong Answers	Category – An (total)	Category – Bn (total)	Accuracy %
Anatomy	20	10	1 (8)	9 (22)	66.6
Biochemistry	26	4	3 (16)	1 (14)	86.6
Community Medicine	20	10	5 (10)	5 (20)	66.6
Forensic Medicine	18	12	0 (0)	12 (30)	60.0
Microbiology	27	3	1 (16)	2 (14)	90.0
Pathology	29	1	1 (29)	0 (1)	96.6
Pharmacology	26	4	3 (15)	1 (15)	86.6
Physiology	21	9	5 (14)	4 (16)	70.0
Total =	**187**	**53**	**19 (108)**	**34 (132)**	**77.9**

Table 2. Overview of Total and Categorical Validation from Re-scored Analysis

Subject	Total Questions*	Correct Answers	Wrong Answers	Category – An (total)	Category – Bn (total)	Accuracy %
Anatomy	10	1	9	0 (1)	9 (9)	10.0
Biochemistry	4	1	3	2 (3)	1 (1)	25.0
Community Medicine	10	3	7	3 (5)	4 (5)	30.0
Forensic Medicine	12	1	11	0 (0)	11 (12)	8.3
Microbiology	3	2	1	0 (1)	1 (2)	66.6
Pathology	1	0	1	1 (1)	0 (0)	0.0
Pharmacology	4	3	1	0 (3)	1 (1)	75.0
Physiology	9	1	8	4 (5)	4 (4)	11.1
Total =	**53**	**12**	**41**	**10 (19)**	**31 (34)**	**22.6**

Legend: Total questions refer to the questions that were answered incorrectly in first (initial) attempt

ISSN 2076-6327

This journal is published by Pitt Open Library Publishing