



@physicaltherapyresearch

RAPID RESEARCH

September 2023

Inside This Week: AI & Healthcare

- ✓ Accuracy of AI in Medical Responses
- ✓ Can ChatGPT Solve Higher Order Problems in Pathology?
- ✓ Can ChatGPT Pass the USA Medical Licensing Examination?

ACCURACY OF AI IN MEDICAL RESPONSES

SEPTEMBER 2023

[Click for Full Text
\(Johnson et al. 2023\)](#)

This pre-print (Not yet peer-reviewed) study evaluated the accuracy and comprehensiveness of ChatGPT-generated responses to medical queries developed by physicians.



KEY FINDINGS

33 physicians across 17 specialties generated 284 medical questions, and graded ChatGPT-generated answers for accuracy and completeness.

Accuracy and Completeness Ratings:

Across all questions:

Median accuracy [5.5/6]

Mean score [4.8/6]

Median completeness [3/6]

Mean score of [2.5/6]

Questions rated easy, medium, and hard:

Median accuracy scores were 6, 5.5, and 5, respectively.

Accuracy scores for binary and descriptive questions (median 6 vs. 5; mean 4.9 vs. 4.7)

Questions with scores of 1-2:

34 were re-queried/re-graded 8-17 days later with substantial improvement (median 2 vs. 4).

MAIN TAKEAWAYS

Integrating language models, such as ChatGPT in medical practice reveals early promise, but with many considerations for safe and optimal use.

While the AI-generated answers displayed high accuracy and completeness scores across various specialties, question types, and difficulty levels, further development is needed to improve the reliability and robustness of these tools before clinical integration.

Medical professionals and patients should be aware of the limitations and actively check AI-generated medical information with trusted sources.

CAN CHATGPT SOLVE HIGHER ORDER PROBLEMS IN PATHOLOGY?

[Click for Full Text
\(Sinha et al. 2023\)](#)

This study aimed to ascertain the capability of ChatGPT in solving higher-order reasoning in the subject of pathology



ChatGPT was used to converse with 100 higher-order reasoning queries. 3 expert pathologists evaluated results on a 0-5 scale & tested accuracy.

Results:

All 100 questions were solved in an average of 45.31 ± 7.14 sec

Median accuracy score was 4.08/5.

Majority (86%) of the responses were "relational" in category.

No difference in the scores of the responses for questions asked from various organ systems in the subject of Pathology.

Scores rated by 3 pathologists had excellent inter-rater reliability (ICC = 0.975)

MAIN TAKEAWAYS

The capability of ChatGPT to solve higher-order reasoning questions in pathology had a relational level of accuracy.

Text output had connections among its parts to provide a meaningful response.

The answers from the program can score approximately 80%.

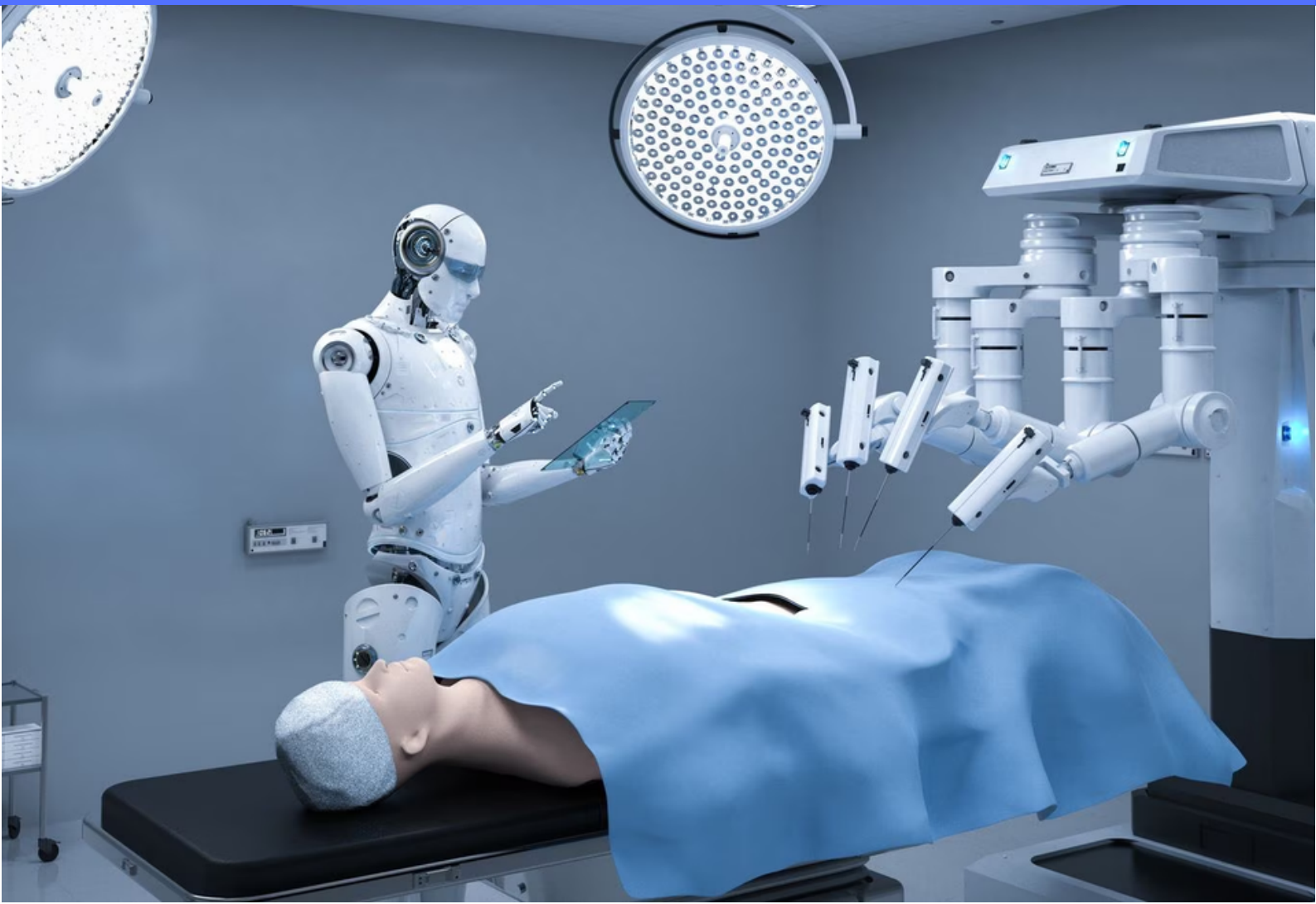
Academicians or students can get help from the program for solving reasoning-type questions also.

As the program is evolving, further studies are needed to find its accuracy level in any further versions.

CAN CHATGPT PASS THE USA MEDICAL LICENSING EXAMINATION?

[Click for Full Text](#)
([Wilson et al. 2023](#))

This study evaluated the performance of ChatGPT on questions within the scope of the United States Medical Licensing Examination Step 1 and Step 2 exams, as well as to analyze responses for user interpretability.



KEY FINDINGS

ChatGPT's performance evaluated on 2 sets of question banks (AMBOSS & NBME) & compared to GPT-3 and InstructGPT.

Accuracy of Answers:

AMBOSS: Step 1 [44%] Step 2 [42%]

NBME: Step 1 [64%] NBME Step 2 [57%]

ChatGPT outperformed InstructGPT by 8.15% across all data sets

GPT-3 performed similarly to random chance.

We found that logical justification for ChatGPT's answer selection was present in 100% of outputs of the NBME data sets.

Internal information to the question was present in 96.8% (183/189) of all questions.

MAIN TAKEAWAYS

ChatGPT performed at a greater than 60% threshold on the NBME-Free-Step-1 data set, achieving the equivalent of a passing score for a third-year medical student.

ChatGPT's showcased a capacity to provide logic and informational context across the majority of answers.

These facts taken together make a compelling case for the potential applications of ChatGPT as an interactive medical education tool to support learning.

GIVE US YOUR FEEDBACK!

MEMBERS

We are on a mission to make research more accessible, easier to interpret, and quicker to implement.

Help us by giving 1 minute of your time to leave feedback for us.

We would greatly appreciate any feedback you have, as it helps us continually improve!

[Leave Review](#)

