

MediSearch Performance Evaluation

1st of August 2023

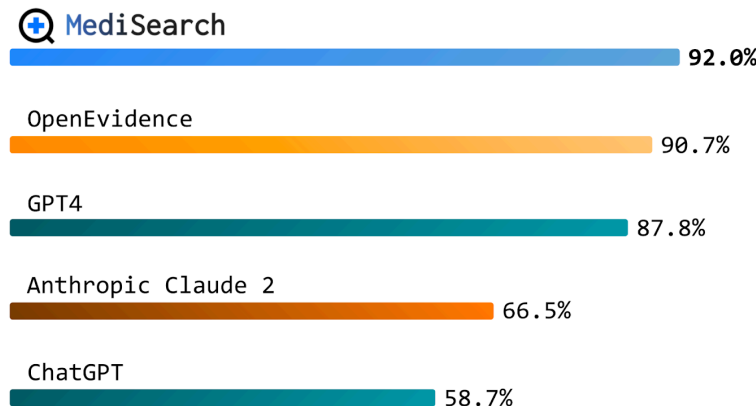
MediSearch is a medical question-answering system. It is available at:

<https://medisearch.io>

The following information is up-to-date as of 1st of August 2023.

USMLE sample exam

MediSearch has a performance of **92%** on the United States medical licensing sample exam (USMLE, <https://www.usmle.org/>, [dataset](#)). You can find the answers that MediSearch produced here: [USMLE Step 1](#), [USMLE Step 2](#), [USMLE Step 3](#). To the best of our knowledge, this is the highest performance out of any question-answering system evaluated on the USMLE sample exam. The following graph illustrates how MediSearch compares to other publically-available models:



The 2022 USMLE sample benchmark was first used to evaluate the medical question answering ability of ChatGPT [1]. We obtained other systems' performances (OpenEvidence [2], GPT4 [3], Claude 2 [4]) from their associated papers and reports.

MedQA

[MedQA](#) is a benchmark similar to the USMLE sample exam. The MedQA dataset is collected from different medical board exams and contains multiple-choice questions to test subjects such

as Internal Medicine, Pediatrics, Psychiatry, and Surgery, among others. We evaluated MediSearch on MedQA's 4-option English test set, which consists of 1,273 questions.

MediSearch achieved a competitive performance of **83.3%**. On the [official leaderboard](#), our result would place MediSearch second, closely behind [Google's Med-PaLM 2](#) (**85.4%**, first place) and in front of Google's [Flan-PaLM](#) (**67.6%**). Based on the MedQA results, MediSearch is the best-performing medical question-answering system available for public use.

You can find the answers that MediSearch produced on MedQA [here](#).