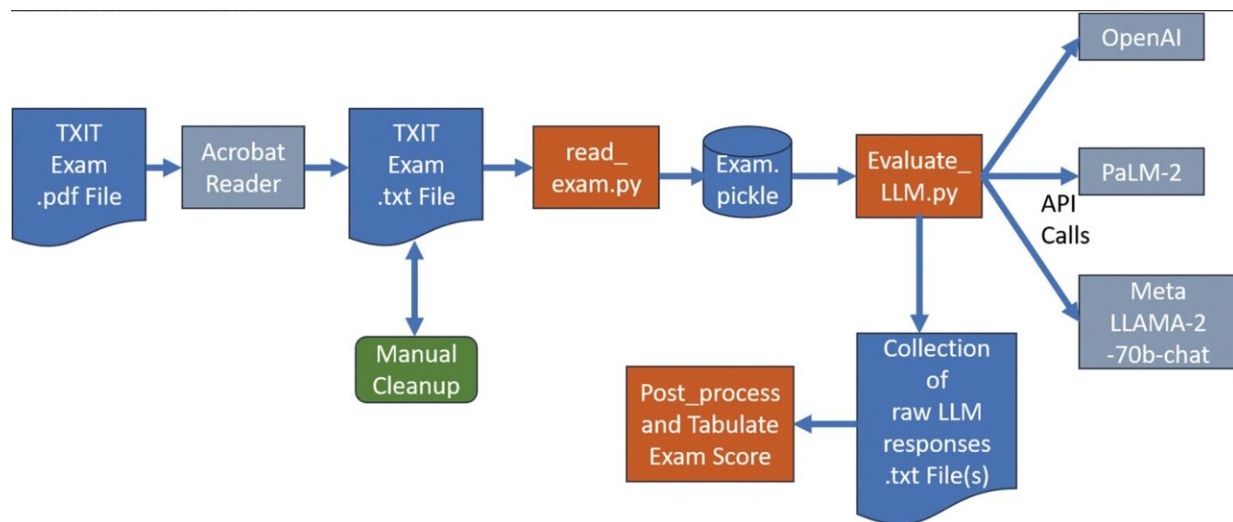


Evaluating the performance of AI-based large language models in radiation oncology

February 8 2024



Architecture of the processing pipeline for evaluation of the 2021 ACR in-training examination with various LLMs. ACR, American College of Radiology; LLMs, large language models. Credit: *AI in Precision Oncology* (2024). DOI: 10.1089/aipo.2023.0007

In a new study [published](#) in the journal *AI in Precision Oncology*, Nikhil Thaker, from Capital Health and Bayta Systems, and co-authors, evaluated the performance of various LLMs, including OpenAI's GPT-3.5-turbo, GPT-4, GPT-4-turbo, Meta's Llama-2 models, and Google's PaLM-2-text-bison. The LLMs were given an exam including 300 questions, and the answers were compared to Radiation Oncology

trainee performance.

The results showed that OpenAI's GPT-4-turbo had the best performance, with 74.2% correct answers, and all three Llama-2 models under-performed. The LLMs tended to excel in the area of statistics, but to underperform in clinical areas, with the exception of GPT-turbo, which performed comparably to upper-level radiation oncology trainees and superiorly to lower-level trainees.

"Future research will need to evaluate the performance of models that are fine-tune trained in [clinical oncology](#)," concluded the investigators. "This study also underscores the need for rigorous validation of LLM-generated information against established medical literature and expert consensus, necessitating expert oversight in their application in [medical education](#) and practice."

"The study highlights the potential of generative AI to revolutionize radiation oncology education and practice. OpenAI's GPT-4-turbo demonstrates that AI can complement [medical training](#), suggesting a future where AI aids in improving patient outcomes. It's essential, though, to validate these technologies rigorously and involve experts to ensure their reliable and effective use in health care," says Douglas Flora, MD, Editor-in-Chief of AI in *Precision Oncology*.

More information: Nikhil G. Thaker et al, Large Language Models Encode Radiation Oncology Domain Knowledge: Performance on the American College of Radiology Standardized Examination, *AI in Precision Oncology* (2024). [DOI: 10.1089/aipo.2023.0007](https://doi.org/10.1089/aipo.2023.0007)

Provided by Mary Ann Liebert, Inc

Citation: Evaluating the performance of AI-based large language models in radiation oncology (2024, February 8) retrieved 29 April 2024 from <https://medicalxpress.com/news/2024-02-ai-based-large-language-oncology.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.