Review of: "Creating a Biomedical Knowledge Base by Addressing GPT's Inaccurate Responses and Benchmarking Context"

Fadi Aljamaan[1]

1 critical care, King Saud University, Saudi Arabia

Potential competing interests: No potential competing interests to declare.

Thanks, authors, for this extensive work.

I do not understand well the objective of your work. Is it to evaluate the GNQA output? Is it to evaluate the quality of the questions generated by experts versus citizen scientists versus GPT?

Regarding the criteria to evaluate the output, how was it set by the RAG for the items you listed? It seems to be broadly subjective criteria?

You did not use a control, like an expert, for example, to compare the evaluations for the query and its answers given by the RAG.

The paper does not run smoothly as you do not separate methods from results.

I suggest extensive revision of the paper by separating the methods from results and explaining the methods in a smoother way.

For the objective and evaluation criteria, they need clarification.