

Dear AI4E Committee,

We have addressed the points shown in the reviews carefully.

1. “The technical details of the paper (after the Introduction) are not written in a self-sufficient manner.” We considered that this problem is caused by very short work. In order to match the maximum 6 pages limitation, we compressed the related works. But now, we have put all important techniques and information back to the paper. Including the introduction about using LLMs for synthetic data generation, medical questions reasoning, and our previous work “NoteAid”.
2. “Table 2 should have a longer caption, explaining in more detail its content.” We added two explanations related to table examples in order to help readers understand what this statistic did.
3. “Manual validation is on a too-small scale. Why the manual validation was not conducted with more samples? Related to that, it is not clearly explained why the LLM evaluation was good enough for the first step in evaluation.” Due to the limitation of the resources, our manual evaluation set is small. And it is hard for us to find more experts to do the evaluation during this period. But getting more manual evaluation is included in our future plan. Besides, we provided reasons for using LLM evaluations by adding related works about its medical question reasoning ability.
4. “There are far too many typos, and incorrect language sentences.” We have done the proofreading carefully to ensure no more typos and errors occurred.
5. “Seemingly 2100 instances for the former, and 84 instances for the latter? (Though this is not completely clear).” We specified the number of evaluations in the paper. This time readers won’t get confused with our experiments and evaluations.

Thank you for your consideration for this paper.

Sincerely,

Xiaocheng Zhang, Master

xiaochengzha@umass.edu

Zonghai Yao, PhD

zonghaiyao@umass.edu

University of Massachusetts Amherst

Hong Yu, Professor

Hong_Yu@uml.edu

University of Massachusetts Lowell