Qeios

Peer Review

# Review of: "Fine-Tuning PHI-3 for Multiple-Choice Question Answering: Methodology, Results, and Challenges"

**Esperanza Johnson**[1]

1. Independent researcher

Interesting piece of work. The author does a nice job of showing the different approaches of LLMs and the specific issue they are trying to tackle in this paper. They also show a flowchart of the methodology followed from the dataset selection until the fine-tuning of the model. The issue of overfitting was stated but not elaborated on, as the author states that “this observation necessitated revisions to the prompt design, as detailed in Section 4.2,” but this is either a typo (referring to section 5.2) or a mistake of not adding all the content. In what would be Section 5 (Results and Discussion), the limitations of overfitting were stated, but the more specific approach to revisions of the prompt design was not elaborated on, outside of Section 3.2, where it states the modification using Alpaca-style prompts. The author should revise this and elaborate further on the prompt design issue. Given the importance of fine-tuning in the final results, more details should be added to the process to better show the more detailed steps that took the accuracy from 62% to 90.8%, which is a substantial jump. Details would enable others to replicate the experiment as best as possible with the same dataset and parameters. Also, the code is not accessible, though that might be given that it is a pre-print at the moment.

## Declarations

**Potential competing interests:** No potential competing interests to declare.



doi.org/10.32388/XBW6K0