

Agency Name:

National Science Foundation

Agency Tracking Number:

2451691

Panel Summary

Description of project:

This proposal aims to improve mathematical problem-solving abilities by proposing hybrid AI systems that integrate large language models (LLMs) with symbolic computation tools, such as Wolfram Alpha, Mathematica, SymPy, etc. To this end, the proposal plans to explore model architecture, task switching, mathematical expression parsing, multimodal fusion, and evaluation and benchmarks. The proposal also describes various aspects of broader impacts on science, engineering, and industry.

Intellectual Merit:

Strengths:

- + The proposed topic of integrating LLMs with symbolic computation tools is interesting, relevant, and important.
- + The experiment plan is well-organized.

Weaknesses:

- The proposal lacks technical content and technical novelty. In particular, there is a lack of discussions on the novel methodology that may help achieve improved performance in mathematical problem-solving.
- Since LLMs are a fast-moving field in both academia and industry, and since it is well-known that commercial LLMs' math capabilities are consistently improving, it is unclear whether the proposed limitations of LLMs will still be valid or significant when the PI completes the preliminary exploration of this topic.
- LLMs are well-known to be extremely computation-demanding. Based on the proposal, it is unclear and unconvincing that the PI has sufficient computation resources to achieve the proposed goal, which is to outperform commercial LLMs.

Broader Impacts:

Strengths:

- + The proposal discusses various aspects of its broader impacts on academia, industry, and engineering.

Weaknesses:

- There is a lack of discussions on education plans and other outreach plans.

Solicitation-specific Review Criteria:

1. There is potential to produce some preliminary results to serve as basis for future competitive research proposals. But the panel has concerns over whether the PI has sufficient resources to run large-scale experiments or whether the PI can achieve the proposed goals before the commercial systems.
2. This proposal can serve as a starting point for this line of research.

Data management plan:

Sufficient.

Mentoring plan for grad students and postdocs:

Good.

Constructive suggestions for improvement:

- For research that requires significant computation resources like LLMs, it can be helpful to explore industrial collaboration opportunities since industry has far greater computation resources than academia could afford, especially for junior faculty.
- NSF also provides opportunities for computation resources, see <https://new.nsf.gov/focus-areas/artificial-intelligence/nairr> and <https://nairrpilot.org/>
- Instead of directly competing with the industry, the panel suggests finding areas that are important but unlikely to be explored in the industry.
- More technical details should be provided in Section 3, including what methods the proposal plans to propose or adopt and the technical explanation of these methods.
- Section 7 should be moved to the Budget Justification document.
- The broader impacts are suggested to discuss education plans and outreach plans, e.g. developing new courses related to the proposed research, K-12 activities to promote STEM participation from underrepresented groups, etc.

Summary Statement: This proposal aims to improve mathematical problem-solving abilities by proposing hybrid AI systems that integrate large language models (LLMs) with symbolic computation tools. Though the panel agrees that the proposed idea is interesting, there are major concerns about the lack of technical novelty, the insufficient resources, and whether the proposed research can outperform commercial systems.

The summary was read by/to the panel and the panel concurred that the summary accurately reflects the panel discussion.

PANEL RECOMMENDATION: Non Competitive