Intellectual Merit Criterion

Overall Assessment of Intellectual Merit

Very Good

Explanation to Applicant

The applicant presents a comprehensive analysis of leveraging large language models for General Visual Question Answering (VQA) with a modular approach. By orchestrating sub-skills, it tackles complex tasks, addressing alignment issues prevalent in end-to-end VQA systems. The proposed use of Retrieval Augmented Generation (RAG) to fill knowledge gaps in text-based queries and connect visual data to external knowledge bases enhances domain-specific VQA. The intellectual merit lies in the exploration of multimodality alternatives, contributing valuable insights for adaptable system development. Testing on established benchmarks and addressing challenges in image embeddings and retrieval processes demonstrates a commitment to practical implementation and problem-solving. Overall, the proposal showcases intellectual depth and innovation in tackling real-world AI challenges.

Broader Impacts Criterion

Overall Assessment of Broader Impacts

Very Good

Explanation to Applicant

The applicant outlines a comprehensive strategy for a Large Language Model Agent in VQA, emphasizing modularity and sub-skills. The proposed approach, though potentially less adept at complex conversational reasoning, aims for broader object generalization and real-world applicability. The enhancement of domain-specific VQA through Retrieval Augmented Generation demonstrates a commitment to addressing knowledge gaps, especially in visual contexts. The broader impact envisions significant strides in Visual AI, benefiting visually impaired individuals and revolutionizing video comprehension for applications in robotics and live video stream management. This ambitious initiative showcases potential breakthroughs with societal implications.

Summary Comments

The applicant possesses a remarkable range of achievements spanning various domains of study and research. These accomplishments encompass publications, scholarships, accolades, and presentations. Notably, the applicant exhibits a wide-ranging interest in both artificial intelligence and natural language processing disciplines. Furthermore, the applicant demonstrates a strong dedication to innovation, applying theoretical knowledge to practical situations, solving real-world problems, and generating insights that can contribute to the evolution of adaptable systems in the future.

Intellectual Merit Criterion

Overall Assessment of Intellectual Merit

Very Good

Explanation to Applicant

The applicant's academic background and research experience are evident. The applicant has prior work experience in AI. The applicant has published several research related to the proposed work. The proposal proposes a modular approach to multimodality using large language models. The applicant has the support of the institution which is evident from the letters of recommendation. While the proposal holds promise for contributing significantly to the field, a more detailed discussion on the methodology would strengthen it.

Broader Impacts Criterion

Overall Assessment of Broader Impacts
Very Good

Explanation to Applicant
The applicant has diverse leadership experience and has received several awards. The applicant has participated in numerous outreach initiatives, with a focus on computer science and AI. The applicant has started a machine learning club at the applicant's university and has been involved in many clubs including the ACM SIG AI club. The proposed work has the potential to advance applications in robotics and embodied assistants.

Summary Comments
The student is in good academic standing and has published several undergraduate research papers. The applicant's dedication to computer science and AI outreach has been demonstrated. In general, the applicant is recommended and has shown solid broader impacts.

Intellectual Merit Criterion

Overall Assessment of Intellectual Merit
Very Good

Explanation to Applicant
Video Question Answering is an essential capability of visual AI assistant. The applicant clearly identified that problems with existing approaches and proposed to leverage highly performant computer vision models to make LLM versatile for the visual modality. It is a great idea to enhance LLM with domain-specific knowledge. However, the research execution plan is unclear although the applicant seems to have extensive experience with computer vision model.

Broader Impacts Criterion

Overall Assessment of Broader Impacts
Very Good

Explanation to Applicant
The applicant has been engaging in promoting AI in the community. The proposed research can potentially generate impacts broader than what anticipated in the proposal.

Summary Comments
The applicant is a generation one college student with excellent research experience who is pursuing a PhD training in AI. The proposal demonstrates a clear understanding of the research problem, an important step in conducting research. Video Question Answering techniques have a variety of applications.

Intellectual Merit Criterion

Overall Assessment of Intellectual Merit
Very Good

Explanation to Applicant
The applicant has earned admission to, and generally performed well in, a top CS program in the country, although there were

occasional low grades including in core CS and math courses. He has first-authored two technical reports; although they were not peer-reviewed by academics, one did secure a 250K award from Amazon to support LLM research. His proposal focuses on video question-answering with LLMs. In particular, he aims to improve the robustness of these models to questions involving novel objects and spatio-temporal relationships. It would have been more compelling to see a little more detail on how this would be achieved: how will architecture and training be modified; and why should we expect these modifications to work? The proposal briefly mentions incorporating techniques like curriculum learning and retrieval-augmented generation, but it is not clear how these will help with novel objects and especially spatiotemporal relationships.

Broader Impacts Criterion

Overall Assessment of Broader Impacts
Very Good

Explanation to Applicant
Video question-answering is a type of multi-modal AI, which is an active research area. However, it is a less urgent need for society compared to other areas of AI such as biomedical applications, or safety, privacy and fairness of AI. It would strengthen the proposal to connect the research more directly to key societal issues. In terms of strengths, some letter writers praise the applicant's leadership, as evidenced by founding a very popular ML club that has attracted hundreds of students at his university.

Summary Comments
The applicant is a strong student who has been very active and successful with regards to industry-sponsored AI internships and competitions with monetary awards, and demonstrates some early leadership ability. The proposed research focuses on video question-answering and is overall interesting and well-reasoned, although it would benefit from additional technical details and emphasis on broader impacts for society.

Intellectual Merit Criterion

Overall Assessment of Intellectual Merit
Excellent

Explanation to Applicant
Strengths The applicant proposes to advance the current state of video, questioning and answering system by using modular large language model agents. The applicant mentions two goals that are to be accomplished for this process. In the first goal, there is discussion of creating smaller VQ sub skills that could be used to solve the problem instead of using end to end VQA systems. There is discussion of retrieval augmented generation, which provides a very good solution to tackle domain, specific inquiries. The applicant also provides supportive, intellectual, merit and broader impact statements. Weakness The applicant could mention sequence of tasks that would be accomplished as a part of this research fellowship in graduate school. The proposal could also benefit from explanation about previous work that ties into the proposed objectives.

Broader Impacts Criterion

Overall Assessment of Broader Impacts
Very Good

Explanation to Applicant
The applicant mentions how these systems can be used to improve real time video comprehension, which can advance a lot of disciplines, like robotics.

Summary Comments

The applicant shows tremendous potential to accomplish the research objectives. The applicant has a strong background in AI and machine learning. The applicant also presents nearly comprehensive discussion on the topic being explored. There are some sections that could be refined in the program. For example, there is a discussion on using iterative approach to integrate additional components in the system. However, the extent to which this iterative approach would be used may have its limitations and could be discussed further. With these things taken into consideration, the project is rated as excellent