### ****Research Statement:****

#### ****Introduction****:

I am passionate about advancing the field of computer vision by exploring new paradigms that combine vision and language. My research interests lie at the intersection of machine learning, computer vision, and natural language processing, specifically focusing on how vision-language models can improve the efficiency and generalization of vision tasks.

#### ****Background****:

During my academic career, I have been fortunate enough to work on several projects that merge different aspects of AI. One of my key projects involved developing an evaluation metric for abstractive long-form summarization, which highlights my ability to contribute to both theory and application. I have also been involved in developing a benchmarking framework for vision-language models in object detection, a project that has deepened my understanding of the challenges in multi-modal learning.

#### ****Research Interests****:

My research interests are driven by two primary goals:

1. **Improving Visual Intelligence**: I am particularly focused on improving the robustness and efficiency of computer vision models. By integrating linguistic features and contextual information into vision tasks, I believe we can build models that are not only more accurate but also capable of understanding the world in a more human-like manner.
2. **Expanding AI’s Real-World Applicability**: I am eager to explore how we can apply these advanced techniques to real-world problems, from autonomous driving to healthcare. For example, I envision using vision-language models to assist in complex decision-making tasks, where both visual inputs and contextual knowledge are crucial.

#### ****Methodological Approach****:

I approach my research with a strong foundation in deep learning and probabilistic modeling. I am particularly interested in designing novel architectures that can handle multi-modal data and scale across different tasks. By leveraging large datasets and employing cutting-edge learning techniques like self-supervised learning, I aim to develop models that are not only accurate but also computationally efficient.

#### ****Future Goals****:

Looking ahead, I intend to continue exploring the integration of vision and language, particularly in the realms of multi-modal learning and zero-shot task generalization. I am also keen to investigate the ethical implications of AI, ensuring that models are interpretable, fair, and accessible to all.

#### ****Conclusion****:

My research is driven by a desire to push the boundaries of computer vision and make AI systems more capable and versatile. Through my work, I hope to contribute to the development of intelligent systems that can understand, reason, and interact with the world in a more holistic and intuitive way.

Yazhe Wan

2024.12.13