My research focuses on building reliable and trustworthy visual recognition systems, with an emphasis on out-of-distribution (OOD) generalization, robustness, and multimodal foundation models. Despite rapid progress in computer vision, modern models often rely on closed-world assumptions and can fail unpredictably when deployed in real-world environments. My work is driven by a central question: how can we design vision systems that remain reliable, interpretable, and robust under open-world conditions and distribution shifts?

To address this, I develop principled learning frameworks and algorithms for open-world visual recognition. A major thread of my work studies OOD generalization and detection in vision systems, aiming to enable models to recognize when they encounter novel or shifted data. I have contributed methods that jointly improve generalization and uncertainty awareness, including approaches that leverage human feedback and adaptive data selection to enhance robustness under limited supervision. These contributions aim to move beyond standard benchmark performance toward vision systems that can safely operate in dynamic and evolving environments.

Another key direction of my research lies in understanding and improving multimodal foundation models, particularly vision-language models (VLMs). As vision systems become increasingly integrated with language and reasoning capabilities, ensuring their reliability becomes more challenging. My work investigates how multimodal representations can be disentangled, interpretable, and robust to spurious correlations, as well as how these models behave under distribution shifts and compositional generalization tasks. In parallel, I study the failure modes of large-scale models, including their limitations in reasoning and long-horizon tasks, with the goal of developing more dependable multimodal systems.

I am also interested in vision systems under data scarcity and real-world constraints. In many applications, collecting labeled visual data is expensive, and rare failure cases are difficult to capture. To address this, I explore data-efficient and human-in-the-loop approaches that prioritize informative samples and feedback to improve robustness, calibration, and coverage. This perspective is particularly important for deploying vision systems in safety-critical settings, where reliability matters more than average-case performance.

My research contributions span core computer vision venues such as CVPR, as well as broader machine learning conferences, reflecting an effort to connect vision problems with theoretical and algorithmic foundations. I have also engaged in interdisciplinary collaborations across vision, machine learning, and foundation model research, which has shaped my perspective on building unified frameworks for trustworthy AI systems.

I believe I am a strong candidate for the Doctoral Consortium because my work aligns closely with the evolving challenges in computer vision: moving from closed-world recognition to open-world, multimodal, and reliable perception systems. I am currently at a stage where I am synthesizing my research into a broader vision of trustworthy visual intelligence, and I would greatly benefit from feedback on how to further integrate OOD learning, multimodal representation learning, and foundation model reliability into a cohesive research agenda.

The Doctoral Consortium would provide a valuable opportunity to discuss these ideas with senior researchers and peers, refine my research direction, and gain insights into transitioning to an independent research career. I am particularly interested in feedback on how to design evaluation frameworks and benchmarks that better capture real-world robustness and failure modes in vision systems, as well as how to bridge the gap between theoretical guarantees and practical deployment.

In summary, my research aims to advance reliable and robust computer vision systems that can operate safely in open-world environments. I am excited to contribute to and learn from the CVPR community through the Doctoral Consortium, and to further develop my vision for trustworthy visual intelligence.