Research Statement

My research is centered on advancing reinforcement learning (RL) through specification-guided frameworks that enhance transparency, safety, and interpretability in AI systems. By automating the refinement of RL specifications using human-centric guidance, my work strives to democratize access to sophisticated AI, ensuring that these technologies are both robust and accessible to a diverse range of users.

Under the mentorship of Professor Abhinav Verma, I have developed algorithms that integrate deep RL with user-defined constraints to demystify complex decision-making processes. A key contribution in this area is my work on automating RL specification refinement, presented at PLDI 2024 in collaboration with Professor Đorđe Žikelić. This research not only addresses fundamental challenges in model transparency but also enhances the real-world applicability of RL systems in safety-critical domains.

Complementing this work, I have been involved in the DoD-funded MIXTAPE initiative, where I applied reinforcement learning to generate actionable strategies and enable user-modifiable policy behavior at runtime. This project underscores my commitment to explainable AI by ensuring that RL models remain interpretable even when managing conflicting objectives. Additionally, I am developing an algorithm-agnostic safety shield that scales to higher dimensions through user-friendly constraint modeling. This safety shield is designed to facilitate safe exploration in complex environments, thus bridging the gap between theoretical RL advancements and practical, deployable solutions.

My interdisciplinary approach extends to collaborations with various fields, where I integrate AI techniques into domains such as energy systems optimization and defense applications. These experiences have not only broadened my perspective on the societal impact of AI but have also equipped me with the skills necessary to translate advanced computational methods into tangible benefits for industry and public welfare.

Looking forward, I aim to further refine my work on specification-guided RL by exploring novel methods for integrating probabilistic reasoning and causal inference into autonomous decision-making frameworks. My goal is to develop AI systems that are not only high-performing but also inherently transparent and safe, paving the way for their deployment in critical real-world scenarios. By fostering collaborations across academic and industry boundaries, I seek to contribute to the creation of AI technologies that are both innovative and ethically grounded.

In summary, my research is driven by a commitment to making AI accessible and trustworthy. Through rigorous theoretical development and practical application, I strive to build systems that empower users and transform complex engineering challenges into opportunities for scientific advancement and societal benefit.