## Abstract

Systems utilizing machine learning (ML) and AI are increasingly interacting with more diverse and heterogeneous populations each year. Individuals are algorithmically ranked for job interviews or financial support, and matched with potential partners on dating apps. Nonetheless, many of these current systems utilizing some form of ML or AI have demonstrated systematic unfairness and bias, sub-optimalities, or instabilities. As AI and ML systems become increasingly integrated into our lives, ensuring their fairness and reliability is critical for the public good.

The central goal of my PhD is to create fair and trustworthy AI systems. I will design and proliferate algorithms which not only achieve good performance, but also address fairness, reliability, and stability of both ML predictors and the AI systems utilizing them. To achieve this, I plan on researching two key areas related to trustworthy and fair AI: (1) Algorithmic fairness of systems utilizing machine learned predictors, such as recommender/ranking systems and two-sided marketplaces; and (2) Understanding the practical limits of fairness and robustness with only black-box API model access. Together, my research directions reflect and address both the integration of AI and ML into larger algorithmic systems and the shift towards powerful and closed-source models and LLMs.

**Words**: 200