Through a PhD at Carnegie Mellon University (CMU) machine learning, I aspire to explore the fairness of artificial intelligence (AI) and machine learning (ML) systems, while also researching ways to improve decision-making through AI. Along these lines, my prior research focused on improving the utility and equity of intelligent computing systems and taught me how to perform and present research on AI and ML.

Current Research Motivation: I aim to research whether we can expand the use of AI in fields like medicine and economics, while simultaneously ensuring algorithmic fairness. My interest arose from research I did with Prof. John Dickerson on fairness in ride-pooling and discussions I had with his other students. These discussions exposed me to the broader impacts AI and ML can have on decision-making in fields like medicine and economics. In medicine, techniques like multi-armed bandits (MABs) are used to make treatment decisions for patients, while in economics, deep learning is used to optimize matches in markets like kidney exchange and rideshare. Because of human stochasticity, decisions in medicine and economics are made without perfect information, so AI and ML can be particularly impactful for overcoming this by making approximately optimal decisions. At the same time, working on fairness in rideshare (with Prof. John Dickerson) and NLP (with Prof. Jordan Boyd-Graber) showed me issues regarding the current unreliability of AI and ML systems, specifically concerning bias and equality, creating the need for unbiased decision-making technologies.

Expanding AI and ML Decision Making: My experience working on human-AI collaboration at the MIT Lincoln Labs [1] exposed me to AI-based decision making and demonstrated the current limitations of AI. My project was inspired by recent work [2] which developed AI-based algorithms to assist physicians with diagnosing health conditions from chest cardiographs. Their algorithm partitioned tasks between expert physicians and AI depending on which was more suited for the task. Our goal was to expand their algorithm to take into account the errors of individual physicians through fine-tuning; we believed that if AI could take into account the specific biases of the individual it worked with, then it could deliver more accurate predictions.

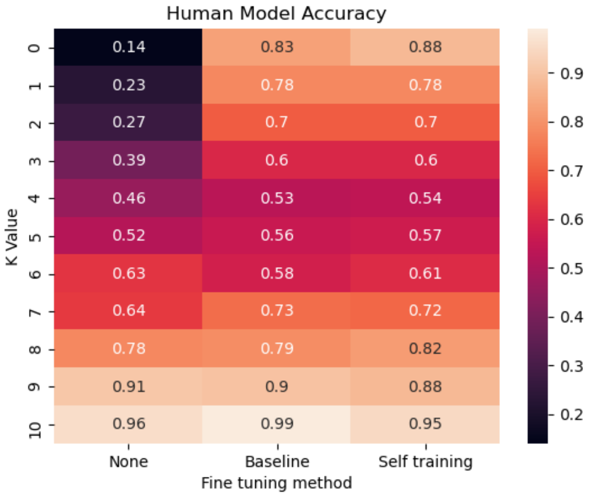

Figure 1: During my work at Lincoln Labs, I found that incorporating fine-tuning for individual-specific data can improve model performance when using synthetic experts.

To do this, I incorporated semi-supervised learning algorithms, like self-training, so partitioning tasks took into account both individual-specific data, and more general unlabeled data from other experts. We found that our algorithm was able to successfully fine-tune only for synthetic experts, whose pattern of skill was simpler, and failed to defer optimally when working with real human experts. Through the project, I learned about issues in AI such as a lack of generalizability and poor decision-making.

To combat these issues, I'm interested in applying robust ML and AI learning algorithms so decisions generalize across situations. This is exemplified by research conducted by **Tuomas Sandholm** [3], which employed ML to predict success rates in heart transplants. Success rate prediction can be helpful when making medical decisions, but determining success rates is difficult due to a variety of stochastic biological factors. To address this, they combined ML models to develop an ensemble that outperformed each model individually. I'm interested in this research because it utilizes the computational power of ML and AI to make decisions, applying these to real-world situations. Other research I'm interested in includes combining game theory with AI, using paradigms such as mechanism design, to make decisions. Game theory can model interactions between agents, which is used with AI to determine approximately optimal decisions while considering agent motivations.

Minimizing Bias: While AI is applicable across a wide range of domains, its use is limited by its potential for unfairness. This issue originates from a variety of sources, including biased data, poor learning algorithms, and bad modeling. While it's important to expand the use of AI, we need to be cautious about avoiding unfairness by analyzing all parts of the model development cycle.

My work on fairness in rideshare [4] and bias in question answering [5] alerted me to the presence of bias in various ML and AI models, while also informing me of methods to combat it. For my project on fairness in rideshare, we found income disparities in driver wages and location-based pickup rate disparities in rider trip requests. To counter this, we developed new fairness-based objective functions, which combined a profit maximization term with variance regularization. One of the toughest parts of the project was defining fairness; for example, on the driver-side, we considered definitions such as the income gap between highest and lowest earners, variance in income, minimum income earned, and other non-income based definitions.

I'm interested in further exploring different notions of fairness, potentially by working with practitioners to determine fairness metrics relevant to their field. An example of this is research conducted by **Steven Wu** that developed fairness metrics for child welfare analysis [6]. To assist developers, I can integrate fairness definitions into model cards, which capture model properties, along with an explanation of why certain metrics are used [7]. Doing so allows practitioners to select models with good performance on relevant fairness metrics.

Future Goals: After graduate school, I plan to become a professor, researching applications of AI and ML, and working with local organizations to ensure the equitable use of AI. My experience as a teaching assistant allowed me to discover my enjoyment of teaching, especially with one-on-one and small-group discussions, so I aim to continue teaching courses and practicing my communication skills at CMU. Becoming a professor allows me to combine my interests in teaching and research.

My research interests fit well with many researchers at CMU. I'm interested in **Tuomas Sandholm's** research on AI and mechanism design, including his work on kidney exchange and auctions. Additionally, **Rayid Ghani's** research lies at the intersection of ML and criminal justice, and ties in nicely with my interest in fairness. My interest in fairness also aligns with research done by **Steven Wu** on making human-facing machine learning more robust, and by **Hoda Heidari** on combining economic modeling with AI/ML for algorithmic fairness. I believe that a PhD at CMU would allow me to produce impactful research that aligns with my research vision.

[1] **Raman, Naveen** and Michael Yee. Improving learning-to-defer algorithms through fine-tuning. *NeurIPS Workshop on Human-Machine Decision Making*, 2021.

[2] Hussein Mozannar and David Sontag. Consistent estimators for learning to defer to an expert. In *International Conference on Machine Learning*, pages 7076–7087. PMLR, 2020.

[3] Brian Ayers, Tuomas Sandholm, Igor Gosev, Sunil Prasad, and Arman Kilic. Using machine learning to improve survival prediction after heart transplantation. *Authorea Preprints*, 2021.

[4] **Raman, Naveen**, Sanket Shah, and John Dickerson. Data-driven methods for balancing fairness and efficiency in ride-pooling. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 363–369. International Joint Conferences on Artificial Intelligence Organization, 8 2021. Main Track.

[5] Andrew Mao, **Raman, Naveen**, Matthew Shu, Eric Li, Franklin Yang, and Jordan Boyd-Graber. Eliciting bias in question answering models through ambiguity. *EMNLP Workshop on Machine Reading for Question Answering*, 2021.

[6] Hao-Fei Cheng, Logan Stapleton, Ruiqi Wang, Paige Bullock, Alexandra Chouldechova, Zhiwei Steven Steven Wu, and Haiyi Zhu. Soliciting stakeholders' fairness notions in child maltreatment predictive systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2021.

[7] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229, 2019.