Ph.D. Statement of Purpose

My research interest is **human-centered artificial intelligence**, lying in the intersection of artificial intelligence (AI) and human-computer interaction (HCI). My experience and research at University of Chicago has inspired my interest in the following problems: (i) generating AI assistance that inspires appropriate trust and reliance on the AI and (ii) using AI to model human perception and intuition.

At UChicago, through Dr. Chenhao Tan's course in Human-Center Machine Learning I learned of the complexities of human-AI interaction. In high-stake domains like medical diagnosis, full automation of AI is often not desired and humans have to make the final decision. I am intrigued by the problems caused by this limitation. For example, for tasks where AI models outperform humans, many AI assistance and explanations improve human performance at the cost of humans blindly following the AI's decision. On the AI side, can we design neutral, unpersuasive AI assistance that retains human agency while improving performance? On the human side, can we train humans to judge AI assistance with more scrutiny. My past research aimed to answer these questions and I plan to continue solving these questions in my future research.

Also, I am motivated by the direct and visible impact in human-centered AI research. My research project [citation] has improved crowdworkers' performance on pneumonia diagnosis in chest X-rays, meaning it has the potential to be implemented in real-life medical settings. Such human-related experiments and improvements are very rewarding to me.

Finally I enjoy the interdisciplinary aspects of human-centered AI research. My past research in human-compatible AI decision-support involved modeling human perception, so we used triplet annotations, a method from experimental psychology. My current research aims to leverage our human-compatible AI build radiology teaching framework and we are exploring psychology literature in learning and categorization. More generally, human-centered AI revolves around how humans interact with a decision-making entity and thus involves many many different fields like economics, sociology, ethics, legal, etc. An exciting aspect of human-centered AI resarch is learning and utilizing knowledge from multiple fields.

At the Pre-Doctoral Masters Program in Computer Science at University of Chicago, I worked with Dr. Chenhao Tan on designing AI-driven decision support and training systems for medical diagnosis. Motivated by AI learning human intuition, we devised a human-compatible model that learned both a classification task and predicting human perception. Such a human-compatible representation learned some form of human similarity function and proved to be useful for case-based decision support.

Our work was a unique project with no preexisting framework to build on, so we face many unprecedented problems. On the data side, we had to deal with issues on the format of human perception data, inter-annotator agreement, the amount of data to collect, etc; on the model side, we had to decide on our model architecture, which layer's embedding to use, the dimension of the embedding, etc.

As human studies were expensive, what helped in forming our decisions were synthetic experiments. I built a synthetic dataset of digitally-generated insects with controllable features and also simulated human agents by tuning weights on the features; I also generated varying decision buondaries by altering the ground-truth labels. Synthetic experiments allowed us to exhaustively explore our design and hyperparameters and also provided more insignht; for example, our method works better for

non-linearly separable distributions.
With positive results from sythntic experiments, we moved on human studies. I conducted a human study on a chest X-ray dataset with Prolific crowdworkers. Compared to an AI that only learned classification, our human-compatible representations provided decision support that led to better performance pneumonia diagnosis.

Of the many research skills I learned, the most important skill to me was using experiments to verify ideas. This was counter-intuitive to me at first since I was used to learning the in-and-outs of a well-rounded idea first and then getting hands-on with it. But our project was a nascent one where we had to build everything from scratch; whenever we had the smallest idea we would run experiments, often using simulated human agents. This eventually led to our ”hunch” that a model learning human perception could be useful in providing decision support.
Another skill is to be flexible and be willing to reshape the problem to be solved. Our initial goal was to design an AI-driven tutorial for radiology residents. In thinking about how to model human learners, we came across literature on AI learning human perception. With many small, incremental ideas and experiments, we found that AI learning human perception produced representations more aligned with humans. But we did not know how this could be useful for teaching, so we detoured to an easier problem of case-based decision support: providing assistance as the test case's nearest-neighbor in the training set using our learned similarity function. After showing our method works for case-based decision support, we return to the problem of teaching humans, with much more foundation to build on.

Our experiment results showed our human-compatible representation leads to better decision support performance than an AI that only learns classification. While the results are positive, many issues remain unsolved in this nascent direction. This research project inspired my interest in the following problems:
(i) generating AI assistance that inspires appropriate trust and reliance on the AI. Many past work provides AI's decision and explanation as assistance and show improved human-AI team performance, but a great part of the improved performance can be attributed to humans simply following AI's suggestion, suggesting overreliance on the AI and a lack of human agency. (citations) Our decision support framework instead aims to provide neutral support that retains as much human agency as possible: we provide example explanations from each class and do not reveal AI's predicted label. However, our experiments also shows providing neutral support leads to lower performance than support that nudges human to follow AI's decision. That is, in human-AI teams there is a tradeoff between human agency and task performance. I am interested in solving this problem as I believe human agency is an important but underexplored issue in human-AI collaboration.

(ii) using AI to learn and model human perception and intuition. As our work showed, AI learning human perception can provide better decisions support performance; I believe in general, incorporating more human knowledge and intution to AI frameworks has potential to provide more effective AI assistance, but many issues remain unsolved in this nascent direction. For example, our work focuses on visual perception on images as it is the easiest form of perception for humans; perception for other modalities like audio and text are also important but much more difficult to collect and utilize. The limitations and bounds of AI learning human perception is also unexplored, including the amount of data needed, the format of data (we used triplet annotations), how well neural networks can learn, etc.

I want to pursue a Ph.D. in human-centered AI because of I enjoy solving problems in human-AI interaction and aspire to do research that has direct social impact. I believe the research at CMU HCII aligns with my interest and motivation.