Ph.D. Statement of Purpose

1   introduction

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 learn and model human perception and intuition.

2   general interest

2.1   interest

My interest in HAI is largely inspired by my research with Dr. Chenhao Tan at the University of Chicago as well as his course in Human-Center Machine Learning. Through course I learned of the human-related side of AI and the numerous complexities in human-AI interaction. When should fully delegate a task to AI who is responsible for the decisions? When AI is used as assistance to humans, how do we design fair and objective assistance instead of merely persuading humans to follow the AI? I am deeply intrigued by the interactions between human and AIs as black-box decision-making entities and I am deeply interested in solving the problems that arise.

2.2   direct social impact

But besides my interest, human-AI interaction and human-centered AI is an important direction of research as AI has increasingly significant influence in humans' decision-making. A core field of human-centered AI lies in high-stake domains like medical or juridical fields. Through my research in this direction, I have potential to produce direct and visible social impact.

2.3   interdisciplinary

Finally I enjoy the interdisciplinary aspects of human-centered AI research. My past research in human-compatible AI decision-support involved modeling human perception, an issue explored by experimental and cognitive psychologists. My current research aims to leverage our human-compatible AI to provide more effective teaching frameworks for radiology residents 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. I greatly enjoy learning and combining knowledge from multiple fields.

3   specific interest

At the Pre-Doctoral Masters program at University of Chicago, I worked with Dr. Chenhao Tan on designing AI-driven decision support and training systems for medical teaching. I have learned many research skills and developed my interests in solving these specific problems: (i) generating AI assistance that inspires appropriate trust and reliance on the AI, (ii) how and when we can achieve human-AI complementary performance, and (iii) using AI to learn and model human perception and intuition.

3.1   project description

Motivated by AI learning human intuition, we devised a human-compatible model that learned both a classification task and predicting human perception. We realized that such a human-compatible representation learned some form of human similarity function and could be leveraged for case-based decision support: providing assistance as the test case's nearest-neighbor in the training set using our learned similarity function. I was responsible for conducting experiments on a synthetic datasets and a human study on a chest X-ray dataset with Prolific crowdworkers. The results showed our human-compatible representation leads to better decision support performance than an AI that onnly learns classification. This research project inspired my interest in the following problems:

3.2   new interests

(i) generating AI assistance that inspires appropriate trust and reliance on the AI An important distinction of our work is that we focus on decision support over mere AI model explanation. Many past works on human-AI team provide AI's decision and some form of explanation as assistance and claim improved team performance, but in most case a great part of the improved performance can be attributed to humans following AI's suggestion. This suggests overreliance on the AI and a lack of human agency, as evident by humans' inability to differentiate AI's errors. 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 work also shows providing neutral support leads to lower performance than providing evidence for model prediction. I believe human-AI teams are inherently human-centered and such neutral supports are desired, but the tradeoff between human agency and task performance is still an significant open problem that I want to solve.

(ii) using AI to learn and model human perception and intuition Our model learns to predict human perception, but in general neural networks learning human perception and intuition is a more difficult task than classification. Perception prediction accuracy is low (70-80%) compared to classification but the exact reason for this is unknown: it could be due to inter-annotation disagreement, randomness in human perception, low-quality annotation, etc. Related work has also primarily focused on visual perception on images and other modalities are less explored. As our work showed, AI learning human perception can provide better decisions support performance; I believe this method has potential to provide more effective AI assistance.

4 my fit with the school/program