Ph.D. Statement of Purpose

Yizhou Harry Tian
University of Chicago

1 short statement

My objective is to pursue a Ph.D. with a focus on **human-centered artificial intelligence**, in the fields of artificial intelligence (AI) and human-computer interaction (HCI). Specifically, I am interested in machine-in-the-loop tasks where human decision-making can benefit from AI assistance while retaining human agency. My past research has been focused on the following 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.

2 general research interest & motivation

2.1 interest in research

My interest in human-centered research can be attributed to the many social science and history courses I took as well as a childhood dream of doing good for mankind. I was never satisfied working solely on the technical, mathematical details of CS; I always aspired tackling issues that were more human-related have more direct and immediate social, "visible" impact. I still remember getting excited when attending a researcher talked about her projects that helped blind children learn in non-blind schools and an interactive projector that helped separated families play games on a table.

2.2 interest in HAI

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. Different from conventional technology that are mostly transparent and in full human control, current AI models are black-box decision-making entities that have may significant influences in humans' decision making, but the influences may not always be positive. In high-stake domains like medical diagnosis, complete reliance on AI decisions may not be desirable. Thus numerous issues arise, like generating AI assistance to inspire appropriate humans trust and reliance on AI, retaining human agency, how to properly explain model decisions, etc. I was intrigued by the plethora of complexities raised by the differences and incompatibilities between humans and AI. I believe human-AI interaction is an important direction of research.

2.3 interest in interdisciplinary research

I also greatly enjoy the interdisciplinary aspects of human-centered AI research. As a Computer Science and Linguistics double major, my initial interest in ML was in NLP as I was excited to connect knowledge from both fields. I fulfilled my interdisciplinary passion as 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 teaching and learning. On a broader scale, 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.

3 research experience

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. Working on this enormous and ambitious project, I have learned many research skills and practices as well as developed my interests in human-centered AI.

3.1 project description

Motivated by AI learning human intuition, we devised a human-compatible model (using a resnet backbone) 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. We conducted experiments on synthetic datasets as well as human studies on a medical dataset and showed our human-compatible representation leads to better decision support performance than a traditional AI representation. (citation)

3.2 challenges

While using AI to learn human perception was not new, our work was the first to combine learning human perception with classification and use the learned representation to assist human decision-making. Thus, there were many unprecedented problems during our work. On the data side, we had to decide on the format of human perception data, inter-annotator agreement, the amount of data, 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. Besides numerous experimental trials, what helped in forming our decisions were synthetic experiments. We build a synthetic dataset of fictional, digitally-generated insects with controllable features and also build synthetic human agents by tuning weights on the features; this allowed to exhaustively explore our design, hyperparameters, and the limitations of our model.

3.3 unique contributions

Working with another Ph.D student, my unique contributions include running synthetic experiments and conducting a human study on a medical dataset. For the synthetic experiment, we build a synthetic dataset of fictional, digitally-generated insects with 2 relevant features and 2 distractor features. By tuning the weights on these 4 features, we build different synthetic human agents. We also generated several datasets with different decision boundaries. Synthetic experiments allowed us to experiment with different hyperparameters, datasets and designs. For the human study, we wrote a survey webapp using Django and deployed it on Prolific. (more description)

3.4 new skills

From this project I gained technical skills. Besides designing and experimenting AI models, including using pytorch and running experiments with Wandb, I also learned of the complications in conducting human studies and interacting with crowdworkers. For example, instructions should be clear and

intuitive. We also needed ways to collect high quality data, so we designed our own attention-checks. I also learned to better visualize and interpret results. (more description)

More importantly, I also learned of soft skills and different perspective of approaching research. Our work of learning human perception for decision support was a unique one and there was no previous work to build on. Thus we went through many peripherally related papers. Through the process I developed lit review skills: the ability to skim through an academic paper and quickly determine its relevance to our research. I also learned to really broaden my scope in lit review and research as work is very interdisciplinary: human perception and decision support are involved in different subfields of psychology, while our framework is applied to the medical field.

Our problem proposal of designing decision-support and training systems is also quite broad and unspecific. Many of our meetings were brainstorming sessions where we would pitch ideas and if one sounded reasonable, we would run simulation experiments on it. This was how we pivoted to learning human perception and joint learning. In short, I learned to not be afraid to throw in ideas and try them out; this sounds trivial but it not easy for me as I have a heavy internal filter and often question myself.

3.5 new interests

This research project inspired my interest in the following problems:

(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) how and when we can achieve human-AI complementary performance. Complementary performance refers to human-AI team performance outperforming human or AI alone. As mentioned, one factor that affects team performance is human agency and reliance on AI. I am also interested in other factors such as human's and AI expertise on the task. I want to explore the limits of human-AI team and when complementary performance is possible.

(iii) 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 problem has potential to provide more effective AI assistance.

4 my fit with the school/program