Han Liu

https://HanLiuAI.github.io hanliu@uchicago.edu

Research Interests

My research focuses on human-AI collaboration. I build AI systems and design novel interactions to assist and teach humans on challenging prediction tasks. My work empower humans with AI by deriving useful explanations from AI and making AI more aligned with humans.

Education

University of Chicago September 2020 - June 2024
Ph.D. in Computer Science

University of Colorado Boulder (transferred out) August 2019 - July 2020
Ph.D. Student in Computer Science

Washington University in St. Louis August 2015 - May 2019
B.A. in Mathematics, Computer Science, and Minor in Linguistics

Research Experience

University of Chicago, Chicago Human+AI Lab September 2020 - Present
Research Assistant (Advisor: Prof. Chenhao Tan) *Chicago, IL*

Project 1: AI-driven Tutorials for Humans on Challenging Image Classification Tasks

- Built deep learning models that learn from both task supervision and human perception signals.
- Developed novel case-based reasoning interactions to provide decision supports for humans in both natural and medical image classification tasks, increasing non-experts' accuracies significantly. [2]
- Designed and implemented teaching algorithms to select important concepts and training examples for fine-grained image classification tasks, achieving competitive performance through simulations.

Project 2: AI-assisted Decision Making for Prostate Cancer Diagnosis

- Built deep learning models that learn from biparametric Magnetic Resonance Imaging (bpMRI) data to detect clinically significant prostate cancer (csPCa), ranked 6th in the PICAI grand challenge.
- Designed and Develped protocols of AI interactions to assist radiologists in prostate cancer diagnosis. Project presented as an educational exhibit at Annual Meeting of the Radiological Society of North America (RSNA 2023). Studies results are in submission.

Project 3: Improving Reinforcement Learning from Human Preferences (RLHF) Algorithms

- Investigated limitations of previous reinforcement learning from human preferences (RLHF) algorithms under conditions of heterogeneous human preference data.
- Extended current methods by generalizing the direct preference optimization algorithm with diverse divergence constraints. [1]

Microsoft Research, Human-AI eXperiences (HAX) Team June 2022 - September 2022
Research Intern (Manager: Dr. Saleema Amershi) *Redmond, WA*

- Designed and piloted user studies of code generation models with multiple human evaluation metrics.
- Conducted analysis on whether offline automatic evaluation metrics align with human values and how they may affect development and deployment decisions. [3]

University of Colorado Boulder, NLP+CSS Lab August 2019 - August 2020
Research Assistant (Advisor: Prof. Chenhao Tan) *Boulder, CO*

- Studied how human and AI collaborate and complement each other under the effect of distribution shift and interactive interfaces in various decision making tasks such as deceptive review detection, profession classification, and recidivism prediction. [4]
- Conducted analysis for large-scale human experiments to study how different types of model-driven tutorials and real-time assistance from model explanations help humans in decision making tasks. [5]

Selected Publications

[1] Chaoqi Wang, Yibo Jiang, Chenghao Yan, **Han Liu**, and Yuxin Chen. Beyond Reverse KL: Generalizing Direct Preference Optimization with Diverse Divergence Constraints. In *International Conference on Learning Representations*, (ICLR 2024, Spotlight).

[2] **Han Liu**, Yizhou Tian, Chacha Chen, Shi Feng, Yuxin Chen, and Chenhao Tan. Learning Human-Compatible Representations for Case-Based Decision Support. In *International Conference on Learning Representations*, (ICLR 2023).

[3] Victor Dibia, Adam Fourney, Gagan Bansal, Forough Poursabzi-Sangdeh, **Han Liu**, and Saleema Amershi. Aligning Offline Metrics and Human Judgments of Value of AI-Pair Programmers. In *Findings of the Association for Computational Linguistics: ACL 2023*, (Findings of ACL 2023).

[4] **Han Liu**, Vivian Lai, and Chenhao Tan. Understanding the Effect of Out-of-distribution Examples and Interactive Explanations on Human-AI Decision Making. *Proceedings of the ACM on Human-Computer Interaction, Volume 5, Issue CSCW2*, (CSCW 2021).

[5] Vivian Lai, **Han Liu**, and Chenhao Tan. "Why is 'Chicago' deceptive?" Towards Building Model-Driven Tutorials for Humans. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, (CHI 2020).

More publications can be found on Google Scholar.

Honors & Awards

- Ranked 6th Place in the PI-CAI Prostate Cancer AI Imaging Grand Challenge November 2022
- NAACL Scholarship for The Undergraduate Summer School at The 2019 Annual Jelinek Memorial Workshop On Speech And Language Technology (JSALT) Top 12.5%, Summer 2019

Professional Service

Conference reviewers for EMNLP, ICWSM, CSCW, FAccT, ICML, and NeurIPS. 2019 - Present.

Teaching Experience

Teaching assistant for the following courses (selected):

CMSC 25100: Introduction to Machine Learning (UChicago, Winter 2024)
CMSC 35400: Machine Learning (UChicago, Sprint 2024)
CSCI 5622: Machine Learning (CU Boulder, Fall 2019)
CSE 559A: Computer Vision (WUSTL, Fall 2018)
CSE 511A: Introduction to Artificial Intelligence (WUSTL, Fall 2018)