

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 Ph.D. Candidate in Computer Science	September 2020 - August 2024
University of Colorado Boulder (transferred out) Ph.D. Student in Computer Science	August 2019 - July 2020
Washington University in St. Louis B.A. in Mathematics, Computer Science, and Minor in Linguistics	August 2015 - May 2019

Research Experience

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

Project 1: AI-driven Tutorials for Image Classification

- Developed deep learning models integrating task supervision and human perception signals, improving non-expert accuracy by 13-17% with model explanations. [2]
- Designed case-based reasoning interactions, significantly enhancing user decision-making in medical and natural image classification.
- Designed and implemented algorithms to select key concepts and training examples for fine-grained image classification, achieving competitive performance on simulated learners.

Project 2: Prostate Cancer Diagnosis with AI

- Built models for prostate cancer detection from MRI, achieving top 6th rank in the PICA challenge.
- Developed AI interaction protocols, assisting radiologists in diagnosis, presented at RSNA 2023.

Project 3: Improving Value Alignment of Large Language Models

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

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

- Designed and piloted user studies of code generation models, using multiple human evaluation metrics.
- Analyzed alignment between offline automatic evaluation metrics and human values, informing development and deployment decisions. [3]

- 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]
- Analyzed results from 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](#).

Technical Skills

Programming Languages	Python, Java, C++, SQL, JavaScript, HTML, CSS
Machine Learning Tools	PyTorch, Keras, Scikit-learn, Pandas, NumPy, Jupyter
Web Development	React, Node.js, Vue.js, jQuery, Django, Bootstrap, MySQL, SQLite

Experience with training and inference with Large Language Models (LLMs) and Vision-Language Models (VLMs): LLaMA-2, Vicuna, LLaVA, CLIP etc.

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)