Han Liu

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

Research Interests

My research focuses on human-AI collaboration and AI alignment. Using techniques in the field of interactive machine learning and machine teaching, I study how humans and AI can learn from each other and how to empower humans with AI.

Education

University of Chicago September 2020 - Present
Ph.D. student 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 Natural and Medical Image Classification

- Building deep learning models that learn from both task supervision and human perception signals.
- Developed a case-based reasoning interaction protocol and policies to provide decision supports for humans in challenging image classification tasks, increasing human prediction accuracy significantly. [1]
- Designing and developing a novel human-AI interaction protocol, an interactive interface, and an example selection algorithm for teaching humans on challenging image classification tasks.

Project 2: Learning from Human Feedback to Improve Language Models

- Studying how human feedback for large language model outputs can improve smaller language models.
- Developing an algorithm to actively select examples for reinforcement learning from human feedback.

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 and conducted analysis on whether offline automatic evaluations align with human values and how they affect development and deployment decisions. (Submission under review.)

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. [2]
- 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. [3]

Selected Publications

[1] **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).

[2] **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).

[3] 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).

[4] Hyunji Hayley Park, Katherine J. Zhang, Coleman Haley, Kenneth Steimel, **Han Liu**, and Lane Schwartz. Morphology matters: a multilingual language modeling analysis. *Transactions of the Association for Computational Linguistics*, (TACL 2021).

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, and ICML.

Teaching Experience

Teaching assistant for the following courses:

CMSC 15200: Introduction to Computer Science II (UChicago, Fall 2020)
CSCI 5622: Machine Learning (CU Boulder, Fall 2019)
CSE 559A: Computer Vision (WUSTL, Fall 2018)
CSE 511A: Introduction to Artificial Intelligence (WUSTL, Fall 2018)
CSE 247: Data Structures and Algorithms (WUSTL, Fall 2017)