Shawn Im

Graduate Student
University of Wisconsin-Madison

Email: shawnim@cs.wisc.edu

Research Interests

AI Safety, Deep learning theory, Interpretability

Education

University of Wisconsin-Madison
Doctor of Philosophy Computer Science — 2023 -

Massachusetts Institute of Technology
B.Sc in Mathematics, Computer Science — 2019 - 2023

Research Experience

University of Wisconsin-Madison — 2023 -
Advisor: Sharon Li

MIT CSAIL — 2022-2023
Advisors: Yilun Zhou, Jacob Andreas

- Developing an evaluation method for saliency maps for image classification based on the saliency map's ability to improve a user's performance on a task representing a practical use case (e.g. model selection) on various types of datasets

MIT Mathematics — 2022-2023
Advisors: Sungwoo Jeong, Alan Edelman

- Studying the spectral properties of Neural Tangent Kernels using Random Matrix Theory particularly for models outside of the linearized regime

Julia Lab — 2020-2021
Advisors: Chris Rackauckas, Alan Edelman

- Developed models for chemical reactions for batteries and for pollutants using surrogate models and Neural ODEs

Media Lab — 2019-2020
Advisors: Takatoshi Yoshida, Hiroshi Ishii

- Developed a model to classify a person's activity (e.g. walking, spinning) while on top of a floor using force sensors

Industry Experience

Amazon Software Engineer Intern — Summer 2021

- Developed an end-to-end AWS framework to delete user data upon request integrating SNS, Lambda, EMR, S3, API Gateway

Publications

Shawn Im, Yixuan Li. Understanding the Learning Dynamics of Alignment with Human Feedback. In Proceedings of International Conference on Machine Learning (ICML), 2024.

Shawn Im, Jacob Andreas, Yilun Zhou. Evaluating the Utility of Model Explanations for Model Development. NeurIPS Workshop on Attributing Model Behavior at Scale (ATTRIB), 2023.

Activities

Grader, Theory of Probability (18.675) — Fall 2022

Math Learning Center Tutor — Fall 2021

- Primarily taught real analysis and calculus

Workshops

Learning Theory Alliance Workshop — October 2022

Relevant Courses

Mathematical Statistics: Non-asymptotic, Stochastic Calculus, Optimization Methods, Deep Learning Theory, Advanced Natural Language Processing, Eigenvalues of Random Matrices, Theory of Probability, Functions of Complex Variable, Statistical Physics, Advanced Data Structures, Combinatorial Analysis, Seminar in Theoretical Computer Science, Introduction to Machine Learning