Brian R.Y. Huang

branhung@alum.mit.edu · (347) 217-1486

Interests: Robust machine learning, reliable AI deployment

EDUCATION

Masters of Engineering, Massachusetts Institute of Technology 2022-2023
Computer Science & Engineering
Advisor: Aleksander Mądry

Bachelors of Science, Massachusetts Institute of Technology 2018-2022
Major in Mathematics | Major in Computer Science & Engineering

RESEARCH - PUBLICATIONS & PROJECTS

Brian Huang and Joe Kwon. "Does It Know?: Probing and Benchmarking Uncertainty in Language Model Latent Beliefs." *NeurIPS Workshop on Attributing Model Behavior at Scale (ATTRIB)*, 2023.

Brian Huang. "Adversarial Learned Soups: neural network averaging for joint clean and robust performance." Master's thesis, 2023. Advised by Hadi Salman and Aleksander Mądry.

Brian Huang and Marcus Khuri. "On Sufficient Conditions for Trapped Surfaces in Spherically Symmetric Spacetimes." Presented at Siemens Competition, 2017.

Other research (class projects & internships)

"Synthetic Instruction Fine-Tuning and Retrieval for Code Generation."
Original research project for 6.S986 Large Language Models and Beyond. Investigated the combination of synthetic datasets, instruction bootstrapping, and in-context learning for code generation in LLMs. Spring 2023.

"Measuring Monosemanticity via Causal Scrubbing."
Final deliverable at Redwood Research, advised by Adrià Garriga-Alonso. Evaluated associations between individual neurons in GPT-2 and semantic concepts, using causal intervention methods for intervening on neuron activations. Winter 2023.

"Markov Chain Monte Carlo for Cipher Breaking."
Final project for 6.437 Inference and Information. Implemented the Metropolis-Hastings method for text decoding, and experimented with algorithmic optimizations to improve decoding performance and speed. Spring 2022.

TEACHING

6.401/6.481 Intro to Statistical Data Analysis - *Graduate Teaching Assistant* Spring 2023
Led problem set and exam grading; led weekly office hours.

6.867 Advanced Machine Learning - *Graduate Teaching Assistant* Fall 2022
Led recitations and office hours for 150+ graduate-level students. Wrote problem sets; graded problem sets and exams.

6.036 Intro to Machine Learning - *Lab Assistant* Spring 2022
Assisted with twice-weekly interactive lab and quiz sessions.

18.S096 Intro to Mathematical Reasoning - *Teaching Assistant* Spring 2022
Graded all problem sets and exams.

Summer STEM Institute (SSI) - *Research Mentor* Summer 2020
Advised research sprints and technical reading deep-dives for 3 high school mentees across deep learning, data science, and number theory.

SELECTED COURSEWORK

Machine Learning	6.867 Advanced Machine Learning, 6.437 Inference and Information, 6.864 Advanced Natural Language Processing, 6.S986 Large Language Models and Beyond, 6.865 Computational Photography
Math & TCS	18.701 Abstract Algebra, 18.404 Theory of Computation, 6.046 Analysis of Algorithms, 18.821 Math Project Lab, 18.204 Discrete Math Seminar, 18.600 Probability, 18.650 Statistics, 18.211 Combinatorics
Applied CS	6.170 Software Studio, 6.031 Software Construction, 6.033 Computer Systems Engineering

AWARDS

Scholar, Regeneron Science Talent Search 2018

Honorable Mention, Davidson Fellows 2018

National Finalist, Siemens Competition 2017

USA Math Olympiad Qualifier (2x) 2017-2018

WORK EXPERIENCE

Matician - *Research Engineer, Perception* Fall 2023 - Current
Trajectory planning and scene understanding for household robot applications.

Mądry Lab, MIT CSAIL - *Graduate Research Assistant* Spring 2022 - Summer 2023
Developed new architecture and method to fine-tune computer vision models for improvements with respect to adversarial robustness and distribution shift. Benchmarked overall approach across a comprehensive range of models (ResNets, CLIPs, CNNs, ViTs) and datasets (CIFAR10, adversarial attacks, ImageNet domain shifts).
Contributed training optimizations and custom data augmentations to the ffcv open-source library for accelerated computer vision training.

Redwood Research - *Research Resident* Winter 2023
Causal intervention methods for mechanistic interpretability in large language models.

JPMorgan Chase - *Quantitative Research Intern* Summer 2021 & Winter 2021
Developed Python-based data processing pipelines and algorithmic optimizations for options risk measures (hVaR). Reduced computational bottlenecks with vectorization in numpy and pandas.
Implemented user interface for traders to calculate option theoretical values and option greeks.

Centrly - *Software Engineering Intern* Summer 2020
B2B SaaS startup. Launched recommendation system for an industry analytics software; implemented graph theory & NLP methods (graph search, PageRank, cosine similarity).

WorldQuant - *Quantitative Research Intern* Summer 2019
Developed statistical methods in C++ to reduce market impact of equities trading algorithms.

Stony Brook University - *Student Researcher* Spring 2017 - Fall 2017
Advised by Marcus Khuri. Investigated and proved novel differential-geometric properties of the onset of black hole formation in general relativity.

SKILLS & ACTIVITIES

Skills Strong in Rust, Python, PyTorch, numpy, matplotlib, git, LaTeX, Jupyter.
Some experience in Javascript, Java, C++, pandas, numba, JAX, sklearn.

Activities MIT AI Alignment Reading Group (member), spring 2023.
AI@MIT project labs (member), fall 2021.
Sloan Business Club, director of member education program, fall 2019.
MIT Music Production Collaborative, general exec, 2020-2022.
MIT/Wellesley Toons Acapella, spring 2019.