# Brian R.Y. Huang

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

**EDUCATION**

**Massachusetts Institute of Technology (MIT)** — **Cambridge, MA**
B.S. in Mathematics, B.S. in Computer Science & Engineering [GPA 4.8/5.0] — *2018 - 2022*
M.Eng. in Computer Science & Engineering [GPA 4.9/5.0]. Advised by Aleksander Mądry. — *2022 - 2023*

*Selected courses*
CS & ML: Machine Learning (G), Inference and Information (G), Natural Language Processing (G), Large Language Models (G), Theory of Computation, Algorithms, Software Studio, Computer Systems Eng
Math: Abstract Algebra, Probability, Statistics, Combinatorics, Number Theory, Discrete Math Seminar

---

**EXPERIENCE**

**Haize Labs** — **New York, NY**
*Researcher, Contract* — *June 2023 - Current*
- Conducting research on safety and robustness of frontier LLM systems. On **Anthropic** redteaming contract.
- Previously worked on interpretability-based methods for LLM factuality; paper in **NeurIPS workshop '23**.

**Matic Robots,** *Research Engineer (Autonomous Driving)* — *Sep 2023 - May 2023*
- Trained 3D occupancy neural nets with model distillation. Led migration of neural net inference systems to NVIDIA/TensorRT. Built Rust-based visual tools for debugging neural net inference and mapping algorithms.

**MIT CSAIL, Mądry Lab** — **Cambridge, MA**
*Graduate Researcher (Computer Vision, Science of Deep Learning)* — *Feb 2022 - Aug 2023*
- Developed new model architecture and method to finetune computer vision models for adversarial robustness, borrowing from weight-space model ensembles and linear mode connectivity. Benchmarked across a range of models (CNNs, CLIPs, ViTs) and data (CIFAR10, adversarial attacks, ImageNet distribution shifts).
- Contributed *numba*-based optimizations and custom data augmentations to the open-source *ffcv* library for accelerated computer vision training. Obtained ~1.2x speedup on the YOLOv5 object detection model.

**Redwood Research,** *Research Resident* — *Jan 2023*
- Analyzed semantic concept-based interpretability in LLMs using causal intervention techniques on neurons.

**JPMorgan Chase,** *Quantitative Research Intern* — *Jun - Aug 2021, Jan 2021*
- Developed data processing pipelines for options risk measures (VaR) with vectorized numpy/pandas code.
- Implemented user interface for traders to calculate option theoretical values and option greeks.

**WorldQuant,** *Quantitative Research Intern* — *Jun - Aug 2019*
- Researched and backtested statistical methods in C++ to reduce market impact of equities trading algorithms.

**Stony Brook University Mathematics,** *Student Researcher (General Relativity)* — *Jun 2017 - Dec 2017*
- Advised by Marcus Khuri. Studied black hole formation in the curved-spacetime setting of general relativity.
- Used methods from differential geometry and tensor calculus to improve theoretical bounds on the formation of "trapped surfaces", a mathematical precursor to black holes. Presented at **Siemens Competition '17**.

---

**TEACHING & SERVICE**

**MIT EECS & Math** — **Cambridge, MA**
*Graduate Teaching Assistant* — *Feb 2022 - Aug 2023*
- TA for Intro to Statistical Data Analysis (G) in spring 2023; TA for Machine Learning (G) in fall 2022; UTA for Intro to Math Reasoning in spring 2022; Lab Assistant for Intro to Machine Learning in spring 2022.
- Led recitations, office hours for 150+ students; wrote homework problems; graded problem sets and exams.

**Momentum AI @MIT**. Taught introductory AI curriculum to low-income high school students in summer 2023.
**Summer STEM Institute (SSI)**. Mentor and speaker for high school science research program in summer 2020.

---

**PUBLICATIONS & PREPRINTS**

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

"Adversarial Learned Soups: neural network averaging for joint clean and robust performance."
**Brian R.Y. Huang**. Master's thesis. 2023.

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

---

**AWARDS**

**Siemens Competition National Winner, 2017** ($25,000 scholarship, 3rd out of 1800+ research submissions)
**USA Math Olympiad Qualifier (2x) | Regeneron STS Scholar, 2018 | Davidson Fellow HM, 2018**

---

**SKILLS**

*Proficient* in Rust, Python, PyTorch, numpy, git, LaTeX, jupyter. *Capable* in JavaScript, Java, C++, pandas, bash.