Tony T. Wang

twang6@mit.edu
tonytwang.net

Education

Massachusetts Institute of Technology — Cambridge, MA
Ph.D. in Computer Science, advised by Nir Shavit — 2021 – Present

M.Eng. in Computer Science, advised by Gregory W. Wornell, GPA: 5.0/5.0 — 2020 – 2021

B.Sc. Double Major in Computer Science and Math, GPA: 4.9/5.0 — 2016 – 2020

- *Master's thesis:* Adversarial Examples in Simpler Settings.
- *Selected CS coursework:* Machine Learning, Inference and Information, Robotic Manipulation, Formal Reasoning about Programs, Cryptography, Compilers, Performance Engineering, Randomized Algorithms, Quantum Computation.
- *Selected math coursework:* Measure Theoretic Probability, Complex Analysis, Functional Analysis, Differential Geometry, General Relativity, Abstract Algebra.

Publications

Google Scholar: https://scholar.google.com/citations?user=YWiob00AAAAJ&hl=en.

Work and Research Experience

Center for AI Standards and Innovation, NIST — Washington, D.C.
Member of Technical Staff — July 2024 – Present

- Assisting the US Government in its efforts to steer AI development to promote human flourishing, economic competitiveness, and national security.

Shavit Lab, MIT — Cambridge, MA
Research Assistant — Fall 2021 – Present

- Working on AI safety, with a focus on adversarial robustness and interpretability.

Astra Fellowship, Constellation — Berkeley, CA
Research Fellow — Jan 2024 – Jun 2024

- Working on language model jailbreak defense.

Genesis Therapeutics — Burlingame, CA
AI Engineer Intern — Summer 2021

- Worked on deep neural networks for molecular property prediction.

Signals, Information, and Algorithms Laboratory, MIT — Cambridge, MA
Research Assistant (M.Eng.) — Summer 2020 – Spring 2021

- Studied toy examples of adversarial examples to unify different aspects of the phenomenon.
- Collaborated with researchers at the Poggio Lab on neurosymbolic algorithms for solving the Abstraction and Reasoning Corpus.

Nvidia — Santa Clara, CA
AI-Infra Research Intern — Summer 2019

- Researched active learning for self-driving vision models, with a focus on diversity-aware batch-mode sampling.

Five Rings Capital New York City, NY
Quant Research Intern Q1 2019

- Analyzed market data for statistical arbitrage opportunities.

Dropbox San Francisco, CA
Network Reliability Engineering Intern Summer 2018

- Automated traffic draining for production routers.

Awards

Eric and Wendy Schmidt Center PhD Fellowship 2022 - 2023

MIT EECS Harold Hazen Teaching Award 2021

Undergraduate Teaching Assistant Award 2020

USA Computing Olympiad finalist (national top 24) 2013, 2015

Other projects

Codeforces Round #336 Q4 2015
Competitive programming contest

- Main organizer and problem writer.
- Drew 3000+ participants.
- Particularly proud of authoring codeforces.com/contest/607/problem/C.

Other Activities

MIT AI Alignment 2022 – Present
Member, Advisor

MIT Club Tennis 2022 – 2024
Member

MIT Anime Club 2016 – 2021
Member, President, Webmaster

MIT Chamber Music Society 2016 – 2020
Violinist

Peninsula Youth Orchestra 2011 – 2016
Violinist, Assistant Concertmaster