## Tim Hua

Neel Nanda's MATS Mentee. Formerly on John List's economics team at Walmart

Website: <a href="mailto:timhua.me">timhua.me</a> **EDUCATION & SKILLS** 

Middlebury College

Middlebury, VT

BA, summa cum laude, Economics (GPA 3.97/4.00)

May 2023

D.K. Smith Economics Prize for best thesis (Fox News's Effect on Social and Moral Preferences).

• Thesis cited by Harvard economist Benjamin Enke in Moral Boundaries

**Harvard College** 

Cambridge, MA

Visiting Undergraduate Student (GPA: 4.0/4.0)

2021 - 2022 School Year

Coursework in economics and math.

**General skills:** Pytorch/Python, R, SQL, Stata.

### **AI SAFETY**

### Neel Nanda and Samuel Marks's MATS Scholar

Berkeley, CA

Scholar

Apr 2025 - Current

- Working on alignment auditing techniques in Neel Nanda's MATS stream with <u>Andrew Qin</u>. Sam Marks is co-advising our project.
- Training model organisms that act differently in evaluation v. deployment and eliciting the deployment behavior on obviously evaluation contexts through steering.
- Earlier work studied uncovering "semantic backdoors" in models.

### Tyler Tracy's MARS Mentee in AI Control

Berkeley, CA

Researcher

Jan 2025 – June 2025

Jan 2025 – June 2025

- Developed theory and ran experiments on optimally combining runtime monitors with different costs (e.g., probes v. black box monitors).
- See our alignment forum post here or our arXiv paper here for results.

# Alice Rigg's AI Safety Camp Mentee in Mechanistic Interpretability

Remote

• Studied the functional role of SAE error terms. See alignment forum post here.

### Miscellaneous

- BlueDot Economics of Transformative AI Course Facilitator (Feb 2025 Cohort)
- Wrote a critique of the bias in bios experiment in Sparse Feature Circuits on alignment forum.

#### PROFESSIONAL EXPERIENCES

#### **Walmart Economics Team**

New York, NY

Manager, Economist Intern in summer 2022, full time July 2023 – Aug 2024 Worked on monetary incentives for drivers on Walmart's last mile delivery platform.

- Designed and executed numerous experiments, some with multiple treatment arms.
- Built causal ML models using causal random forests and other double ML methods
- Wrote R code that conducts synthetic difference in differences estimation and calculates conformal inference-based confidence in parallel. Code reused by others on my team.
- Presented work to Walmart chief economist John List, <u>fifth ranked economist in the world</u>
- Translated technical findings and frameworks for non-technical partner business teams As an intern, created pricing zones that minimized customers' cross-zone shopping by constructing a weighted graph of Walmart stores and partitioning the graph.
- Pricing zones are still used at Walmart today