Tim Hua

Independent AI Safety Researcher. Formerly on John List's economics team at Walmart

timhua0@gmail.com   Website: timhua.me

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
- Took courses economics theory, probability and statistics, real analysis, etc.

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

AI SAFETY RESEARCH

Neel Nanda MATS Training Program   Remote
Scholar   Apr 2025 - Current
- Accepted (with funding) to Neel's training program in mechanistic interpretability.

Tyler Tracy's MARS Mentee in AI Control   Berkeley, CA
Researcher   Jan 2025 - Current
- Working with Tyler Tracy (Redwood) on evaluating and formalizing new control protocols.
- Modeling how to optimally allocate budget between human auditing, an expensive and accurate monitor, and a cheaper but less accurate monitor.
- Presented poster and gave a lightning talk at Control Conf.

Alice Rigg's AI Safety Camp Mentee in Mechanistic Interpretability   Remote
Researcher   Jan 2025 - Current
- Working with Alice Rigg on mechanistic interpretability studying the functional role of SAE error terms.

Miscellaneous projects
- 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.
- Conceived and explored a control protocol where the untrusted model writes a plan and delegates the execution to a trusted model. Writeup and pilot results here.
- Interpreting obfuscating jailbreaks together with Veniamin Veselovsky, some results are here.

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 ML models using causal random forests and other double ML methods, ran hyperparameter sweeps with out-of-sample off-policy validation following *Treatment Allocation Under Uncertain Costs*
- 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, fifth ranked economist in the world
- 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

Harvard Economics Department   Remote
Course Assistant for Professor Edward Glaeser   June 2022 – Dec 2022
- Wrote sections in the textbook of Econ 1011a: Intermedia Microeconomics: Advanced.

Harvard Kennedy School   Cambridge, MA
Research Assistant for Professor Desmond Ang   March 2022 – July 2022
- Scraped tvtropes.org for an exploratory project investigating stereotypes in films.

Brookings Institution   Remote
Research Assistant for Professor Carol Graham   June 2021 – May 2022
- Coauthor on paper examining mental and physical ill-being during the COVID-19 pandemic.

Middlebury Economics Department   Middlebury, VT
Research Assistant for Professor Marin Abel   September 2020 - March 2021
- Investigate how different compensation schemes could help close the behavior-intention gap.