

# Zihao (Gary) Wang

garywa@umich.edu | Ann Arbor, MI | zihaowang12.com

## EDUCATION

**University of Michigan** | Ann Arbor, MI **Expected Apr 2027**

B.S. in Data Science; Minor in Mathematics | GPA: 3.89 | University Honors (Fall 2025)

**Relevant Coursework:** Information Retrieval, Statistics & AI, Statistical Computing, Database Systems, Data Structures & Algorithms, Probability, Linear Algebra

**University of Rochester** | Rochester, NY **Aug 2023 - May 2025**

Coursework in Data Science, Economics, and Mathematics | GPA: 3.91 | Dean's List (4 semesters)

## TECHNICAL SKILLS

**Programming:** Python, TypeScript, C++, SQL, Java, R

**Libraries/Tools:** React, pandas, NumPy, scikit-learn, Slurm

**Machine Learning:** XGBoost, random forests, sparse autoencoders, vector embeddings

**Analysis:** time-series analysis, geospatial analysis

## EXPERIENCE

**Undergraduate Research Assistant** | University of Michigan **May 2026 - Present**

- Built an end-to-end React/TypeScript prototype for an education-technology research project, developing reusable components and responsive user interfaces.
- Translated evolving research requirements into accessible product workflows while documenting design decisions and incorporating stakeholder feedback.

**Researcher, STEMFORALL** | University of Rochester **Jul 2026 - Present**

- Evaluated sparse-autoencoder training objectives on Pythia-160M activations and identified settings where normalized-MSE training reduced both reconstruction error and next-token loss.
- Prototyped sparse transcoders to replace Pythia-160M feed-forward layers and evaluated behavioral fidelity to the original model.

**Software Engineering Intern** | Beijing 6Cloud Technology **Jun 2025 - Jul 2025**

- Diagnosed 50+ software defects using LLM-generated root-cause hypotheses, reproduction steps, and targeted testing.
- Converted recurring failure patterns into a development and testing checklist spanning design, implementation, and validation.

## PROJECTS

**Social Media Risk Model** **Jan 2026 - Apr 2026**

- Built and evaluated an XGBoost model to estimate self-reported productivity loss from survey data; identified noisy and inconsistent responses as key performance constraints.
- Developed a web application using vector embeddings to match users with similar profiles and visualize comparisons.

**BME Summer Hackathon** | University of Rochester **Aug 2024**

- Led a 4-member team to 7th place among 51 teams by developing a gradient-boosting model that predicted DNA peak-gene associations with 90% precision.

**NYC Transportation Data Science Explorer** | NSDC **May 2024 - Sep 2024**

- Analyzed 1 million NYC collision records using time-series and geospatial methods; designed a severity index and identified the highest-risk ZIP codes.