

Eric Enouen

Email: enouen@cs.cornell.edu | Homepage: ericenouen.github.io

Research Interests

Responsible machine learning, interpretability, human-AI collaboration

Education

Cornell University

August 2024 - May 2029 (Expected)

Ph.D. in Computer Science

The Ohio State University

August 2020 - May 2024

B.S. Computer Science & Engineering, Minor in Mathematics

Research

In Preparation

- **Eric Enouen**, Sainyam Galhotra. Concept Bottleneck Diffusion for Steerable Generation.

Publications

- **Eric Enouen**, Sainyam Galhotra. Debugging Concept Bottleneck Models through Removal and Retraining **ICLR 2026**.
- Jonathan Rosenthal, **Eric Enouen**, Hung Viet Pham, and Lin Tan. DisGUIDE: Disagreement Guided Data-Free Model Extraction. **AAAI 2023, Oral**.
- **Eric Enouen***, Katja Mathesius*, Sean Wang*, Arielle Carr, Sihong Xie. Efficient first-order predictor-corrector multiple objective optimization for fair misinformation detection. **IEEE BigData 2022, Oral**.

* Equal contribution

Awards

- 2025-2026 CIS-Bowers LinkedIn Fellowship (Full tuition/stipend for academic year)
- 2021-2023 NSF REU Fellowship Recipient (Sponsored summer undergraduate research)

Research Experience

Cornell University

August 2024 - Present

Prism Lab

- Advised by [Sainyam Galhotra](#), focusing on how incorporating interpretability into the model design process can provide leverage on broader sociotechnical challenges such as fairness, robustness, and accountability.
- Designed a debugging framework for concept bottleneck models, allowing users to remove undesired concepts and directly align model predictions with human reasoning.

- Worked as part of the Auton Lab with [Artur Dubrawski](#) on prototype-based routing for mixture of experts.
- Created a framework to combine the prediction capabilities of multiple hospitals in an explainable manner, by routing patients to the models most capable of performing well.

- Worked under [Lin Tan](#) exploring data-free model extraction.
- Introduced a novel disagreement loss to generate useful synthetic samples in a query-free way, boosting the extraction performance of prior work by 18.48% on CIFAR-100.

- Worked with [Sihong Xie](#) exploring the accuracy-fairness tradeoff in fair spam detection.
- Utilized a predictor-corrector method to explore tangentially across the optimal trade-off curve, more efficiently finding optimal tradeoffs than prior work.

Skills

Languages: Python, Java, C/C++, C#, MATLAB, x86-64 Assembly
Libraries: PyTorch, TensorFlow, Git

Service

Reviewer

CVPR 2024, 2025; NeurIPS 2025; ICLR 2026

[The Bridge](#) Tutor

September 2022 - May 2024

Help high school students in Columbus City Schools succeed in the classroom and continue on towards higher education. Weekly tutoring sessions during the semester.

Teaching Experience

Teaching Assistant

2024

CS 5780: Introduction to Machine Learning

[CMU RISS RoboLaunch](#) Team

September 2023 - May 2024

On the marketing team for RoboLaunch: introducing undergraduates to broad topics in robotics through talks and interactive workshops.