Ziyao (Richard) Cui

(919)943-1204 | richard.cui@duke.edu | ziyao-cui.github.io

Education

Duke University Durham, NC

Bachelor of Science in Computer Science *Expected May 2026*

- GPA: 4.0
- Majors: Computer Science (Artificial Intelligence and Machine Learning Concentration), Statistical Science (Data Science Concentration), Economics (Minor)
- Advisors: Dr. Jian Pei, Dr. Cynthia Rudin, Dr. Edric Tam
- Honors: The Phi Beta Kappa Honor Society, Dean's List with Distinction
- Courses: Machine Learning Theory & Algorithm, Real Analysis, Bayesian and Modern Statistics

Publications

- **Cui, Ziyao**, Minxing Zhang, and Jian Pei. "On Membership Inference Attacks in Knowledge Distillation." arXiv preprint arXiv:2505.11837 (2025).

Publications In Preparation

- Graph Neural Networks Embeddings
  - **Ziyao Cui**, Edric Tam
- Small Area Estimation with Satellite Imagery
  - **Ziyao Cui**, Sahoko Ishida, Seth Flaxman
- Data Buyer Privacy with Time Series Data (to be submitted to AAAI)
  - **Ziyao Cui**, Minxing Zhang, Jian Pei
- Revisiting Broken Windows Theory (to be submitted to Harvard Data Science Review)
  - **Ziyao Cui**, Erick Jiang, Nicholas Sortisio, Haiyan Wang, Eric Chen, Cynthia Rudin
- Spatio-Temporal Interpolation with Multi-layer Perceptron (to be submitted to ACM Transactions on Spatial Algorithms and Systems)
  - Minxing Zhang, **Ziyao Cui**, Liang Zhao

Research Experience

Small Area Estimation with Satellite Imagery February 2025 – Present

Student Researcher *Oxford, United Kingdom*

- Advisor: Dr. Seth Flaxman, Oxford University
- Designed and implemented a novel Bayesian-oriented Small Area Estimation (SAE) framework using satellite imageries to analyze food security in Africa.
- Leverage computer vision models like DINO to extract features for SAE training.

On Membership Inference Attacks in Knowledge Distillation January 2025 – May 2025

Student Researcher *Durham, NC*

- Advisor: Dr. Jian Pei, Duke University
- Analyzed transferability of Membership Inference Attacks (MIAs) across different Large Language Model (LLM) architectures and dataset domains within the knowledge distillation pipeline, and demonstrated that student models can be more vulnerable to MIAs than teacher models under standard distillation.
- Proposed and implemented five novel privacy-preserving distillation methods to enhance student model privacy, including ensemble-based and soft-label smoothing approaches.

- Led writing and visualization of experimental results, including key tables demonstrating the effectiveness of proposed privacy-preserving distillation methods.

Data Buyer Privacy with Time Series May 2024 – Present

Student Researcher *Durham, NC*

- Advisor: Dr. Jian Pei, Duke University
- Design and implement an attacker model in Python aimed at attacking the data buyer's true purchase intent in a time series given the buyer's privacy-preserving published purchase intents over the time period.
- Leverage Hidden Markov Model and Reinforcement Learning techniques to extract temporal relationships within data buyer's purchase intents and develop holistic learning by taking into account both preceding and succeeding behaviors of the data buyer.
- Adapt the proposed attacker model to predict a person's precise location given a time series of the person's published region.

Revisiting Broken Windows Theory January 2024 – Present

Student Researcher *Durham, NC*

- Advisor: Dr. Cynthia Rudin, Duke University
- Enrolled in Duke course CS 474: Data Science Competition.
- Presented at the American Statistical Association (ASA) Joint Statistical Meetings in August 2024.
- Awarded Honorable Mention in the 2024 ASA Data Expo Challenge Professional Category.
- Explore machine learning techniques such as DBSCAN and PaCMAP to analyze national and local crime datasets, contributing insights into law enforcement strategies.
- Leverage observational causal inference techniques like Matching After Learning to Stretch (MALTS) to reassess Broken Windows Theory, evaluating its validity in actual crime, perceived crime, and inter-city differences by controlling for a significantly larger set of variables than previous works.
- Conducted in-depth variable importance analysis on a national scale and high-crime urban areas, using geospatial crime data to identify key factors influencing crime rates and perceptions.

Spatio-Temporal Interpolation with Multi-layer Perceptron November 2023 – Present

Student Researcher *Durham, NC*

- Advisors: Dr. Liang Zhao, Emory University; Ph.D. Student Minxing Zhang, Duke University
- Develop a framework using a multi-layer perceptron (MLP) in Pytorch with engineered spatio-temporal distance and angle to interpolate targeted index at unknown locations over a time series.
- Design an efficient uncertainty analysis of the MLP approach that takes into account both spatial and temporal aspects.

Duke Impact Investment Group (Duke Student Organization) September 2022 – Present

Data Analyst *Durham, NC*

- Present solutions and key insights for business problems to clients from startups, including Aquatrax, LandUp, and ClassRanked, using data analysis and data mining.
- Employ machine learning techniques such as k-fold cross-validation and Adaboost in Python to classify degraded vs. non-degraded water meters with existing meter readings.
- Develop data pipelines in SQL and Python that extract US soil data from the USDA website to store and query within Google Cloud Platform, supporting future machine learning model developments.

Internship Experience

Outlier AI May 2024 - Present

Coding Expert and Reviewer *Remote*

- Craft and answer questions in Python and Java related to computer science in order to help train AI models.
- Evaluate and rank code generated by AI models to help cutting-edge generative AI models write better code, and review responses by other analysts.

Shanghai Research Institute of Computing Technology Co. May 2024 - August 2024

Research Product Development Intern *Remote*

- Designed and developed feedback models for health monitoring devices.
- Conducted comprehensive research on health monitoring equipment, analyzing data and refining design strategies.
- Gathered insights from over 80 health monitoring devices across 16 domestic and international companies to optimize model performance and enhance device feedback mechanisms.

Teaching Experience

CS 671: Machine Learning Theory & Algorithm (Graduate) August 2024 – December 2024
Undergraduate Teaching Assistant *Durham, NC*

- Plan and lead weekly 50-student discussions of exemplary questions to supplement course material.
- Write and grade homework problems (theory and coding) to assess student understanding on machine learning topics including k-means clustering, boosting, and random forest.
- Host office hours open to 200+ students, assisting with questions from homework and lectures, and answer student questions on Ed discussion forum.
- Meeting weekly with professor and other teaching assistants to discuss course administration issues including verifying homework problems and progress made during discussion sections.

MATH 531: Real Analysis (Graduate) January 2025 – Present
Grader *Durham, NC*

- Evaluated and provided detailed feedback on weekly Real Analysis homework assignments for a graduate-level course with 20+ students, ensuring clarity, rigor, and consistency.

Skills & Interests

Languages: English, Mandarin, Spanish, German
Technical Skills: Python, LaTeX, Java, C, R, SQL, MATLAB, Excel, PowerPoint, Undergraduate Responsible Conduct of Research
Activities: HackDuke, Duke Chronicle, Duke Sports Analytics Club, Carnegie Mellon Sports Analytics Conference
Interests: Photography, Kyokushin Karate, Piano, Skiing, Soccer, Basketball, Boston Celtics
Professional Memberships: American Statistical Association (ASA), Member