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 — *August 2022 - May 2026*

- GPA: 4.0/4.0
- Majors: Computer Science (Artificial Intelligence and Machine Learning Concentration), Statistical Science (Data Science Concentration)
- Minor: Economics
- Research Advisors: Dr. Jian Pei, Dr. Cynthia Rudin, Dr. Edric Tam
- Courses: Machine Learning Theory & Algorithm (Graduate), Real Analysis (Graduate), Data Science Competition, Deep Learning, Optimisation for Data Science, Bayesian and Modern Statistics, Statistical Learning & Inference

University of Oxford — Oxford, United Kingdom
Study Abroad Program in Mathematics and Economics — *Jan 2025 – Jun 2025*

- Research Advisors: Dr. Seth Flaxman, Dr. Sahoko Ishida

PUBLICATIONS

- **Ziyao Cui**, Minxing Zhang, Jian Pei. Learning to Attack: Uncovering Privacy Risks in Sequential Data Releases. *arXiv preprint arXiv:2510.24807*, 2025. *Under review at IEEE International Conference on Data Engineering (ICDE) 2026.*

- **Ziyao Cui**, Edric Tam. Beyond the Laplacian: Interpolated Spectral Augmentation for Graph Neural Networks. *Under review at International Conference on Learning Representations (ICLR) 2026.*

- **Ziyao Cui**, Erick Jiang, Nicholas Sortisio, Haiyan Wang, Eric Chen, and Cynthia Rudin. Revisiting Broken Windows Theory. *arXiv preprint arXiv:2509.16490*, 2025. *Under review at Harvard Data Science Review.*

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

PUBLICATIONS IN PREPARATION

- Sahoko Ishida, Mohammed Aheed Osman, **Ziyao Cui**, Kevin Tang, Seth Flaxman. Bayesian Small Area Estimation of Inadequate Micronutrient Intake. *Expected November 2025; to be submitted to International Journal of Health Geographics.*

- **Ziyao Cui**, Jian Pei. Uncertainty in Knowledge Distillation. *Expected November 2025; to be submitted to ACM KDD 2026.*

- **Ziyao Cui**, Edric Tam. Normalization as a Modeling Choice in Spectral Graph Embeddings. *Expected December 2025; to be submitted to International Conference on Machine Learning (ICML) 2026.*

- **Ziyao Cui**, Sahoko Ishida, Seth Flaxman. FoodSat: Small Area Estimation for Food Security Analysis with Satellite Imagery. *Expected December 2025.*

PROFESSIONAL SERVICE

Reviewer: *NeurIPS* (2025)

HONORS AND AWARDS

Honors: The Phi Beta Kappa Honor Society (Top 1% of Class), Dean's List with Distinction (3 times)

Awards: 2024 ASA Data Expo Challenge Professional Category Honorable Mention

Research Experience

Knowledge Distillation as Uncertainty Transformation
May 2025 – Present

Student Researcher
Durham, NC

- Advisor: Dr. Jian Pei, Duke University
- Conducted theoretical and empirical study on how knowledge distillation reshapes model uncertainty, introducing the concepts of inter-student and intra-student variance and identifying key sources of uncertainties in the knowledge distillation process.
- Demonstrated and proved that distillation systematically suppresses intra-student uncertainty while introducing inter-student uncertainty, revealing tradeoffs between fidelity and expressiveness.
- Proposed solutions to enhance student model uncertainty expectations, including techniques for processing multiple teacher responses for supervision and dividing hallucination responsibilities between teacher and student models.

Interpolated Spectral Augmentation for Graph Neural Networks
May 2025 – September 2025

Student Researcher
Durham, NC

- Advisor: Dr. Edric Tam, Warren Alpert Fellow at Stanford University
- Introduced Interpolated Laplacian Embeddings (ILEs), a new family of spectral embeddings that generalize beyond the Laplacian to capture both community and core–periphery structures in graphs.
- Developed theoretical analysis using spectral graph theory to interpret how different graph operators encode structural information, and provided guidelines for selecting operators under limited or missing node features.
- Demonstrated through simulations and experiments on synthetic (stochastic block models) and real-world datasets (Karate Club, Twitter Congress, Facebook Ego, Polblogs, Cornell, Texas, Wisconsin) that ILE-based augmentation consistently improves GNN performance compared to Laplacian defaults.

Bayesian Small Area Estimation of Inadequate Micronutrient Intake
April 2025 – Present

Student Researcher
Oxford, United Kingdom

- Advisors: Dr. Seth Flaxman and Dr. Sahoko Ishida, Oxford University
- Developed and validated Bayesian small area estimation models to generate high-resolution maps of inadequate iron, folate, and vitamin B12 intake in low- and middle-income countries, enabling policy-relevant nutrition insights at subnational levels.
- Applied both area-level (Fay–Herriot extensions with variance smoothing) and unit-level (Beta-Binomial with spatial random effects) frameworks in a fully Bayesian setting, ensuring rigorous uncertainty quantification.
- Conducted validation with Rwanda's nationally representative EICV7 survey and extended methods to Nigeria's state-level data, demonstrating reliable estimation at finer administrative levels where direct survey data are sparse or unavailable.
- Collaborated with the World Food Programme to inform nutrition policy and program design, contributing methodological advances directly to global food security applications.

Small Area Estimation for Food Security Analysis with Satellite Imagery
February 2025 – Present

Student Researcher
Oxford, United Kingdom

- Advisors: Dr. Seth Flaxman and Dr. Sahoko Ishida, Oxford University
- Developed a Bayesian small area estimation framework for food security analysis, validated using Zimbabwe's ZIMVAC survey with ground-truth Admin2-level data and extended to other African countries.
- Incorporated satellite imagery as auxiliary features to the small area estimation model, applying computer vision models like DINO to extract high-dimensional representations that improved predictive accuracy in data-sparse regions.

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 LLM architectures and dataset domains within the knowledge distillation pipeline, and proved that, contrary to the common intuition, 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.

Uncovering Privacy Risks in Sequential Data Releases
Student Researcher

May 2024 – Present

Durham, NC

- Advisor: Dr. Jian Pei, Duke University
- Demonstrate that existing privacy guarantees in individual data releases can be bypassed when sequences are jointly analyzed, revealing critical vulnerabilities in current privacy-preserving mechanisms.
- Design and implement a novel privacy attack framework in Python, combining Hidden Markov Models and reinforcement learning to exploit temporal dependencies in sequential data releases.
- Evaluated the proposed attack model on 4 synthetic and real-world datasets, achieving significantly higher location inference accuracy than baseline models.

Revisiting Broken Windows Theory
Student Researcher

January 2024 – September 2025

Durham, NC

- Advisor: Dr. Cynthia Rudin, Duke University
- 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 how urban structures influence both violent crime and perceptions of safety in New York City and Chicago, contributing insights into law enforcement strategies.
- Leverage observational causal inference techniques like Matching After Learning to Stretch (MALTS) to reassess Broken Windows Theory, disentangling structural effects from demographic confounders, producing robust tract-level and localized treatment effect estimates, and controlling for a significantly larger set of variables than previous works.
- Demonstrated that abandoned buildings and high-foot-traffic structures, like transit hubs and schools, are associated with elevated crime and heightened perceived danger, with heterogeneous effects across cities and subgroups—emphasizing the need for tailored urban policy interventions.

Duke Impact Investment Group (Duke Student Organization)
Data Analyst

September 2022 – Present

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.

TEACHING EXPERIENCE

CS 671: Machine Learning Theory & Algorithm (Graduate)
Undergraduate Teaching Assistant

August 2024 – December 2024

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, random forest, and neural networks.
- 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)
Undergraduate Teaching Assistant

January 2025 – May 2025

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.

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.

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), Royal Statistical Association (RSS), Institute of Mathematical Statistics (IMS)