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: Theory and Algorithms for Machine Learning (Graduate), Real Analysis (Graduate), Data Science Competition, Bayesian and Modern Statistics, Statistical Learning & Inference, Design and Analysis of Algorithms

University of Oxford   Oxford, United Kingdom
Study Abroad Program in Mathematics and Economics   *January 2025 – June 2025*
- Research Advisor: Dr. Seth Flaxman

PUBLICATIONS

- **Ziyao Cui** and Jian Pei. How Is Uncertainty Propagated in Knowledge Distillation? *arXiv preprint arXiv:2601.18909*, 2026. *Under review at the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2026.*

- **Ziyao Cui**, Minxing Zhang, and Jian Pei. On Membership Inference Attacks in Knowledge Distillation. *arXiv preprint arXiv:2505.11837*, 2025. *Under review at the 64th Annual Meeting of the Association for Computational Linguistics (ACL 2026).*

- **Ziyao Cui** and Edric Tam. Beyond the Laplacian: Interpolated Spectral Augmentation for Graph Neural Networks. *arXiv preprint arXiv:2511.11928*, 2025.

- **Ziyao Cui**, Minxing Zhang, and Jian Pei. Learning to Attack: Uncovering Privacy Risks in Sequential Data Releases. *arXiv preprint arXiv:2510.24807*, 2025.

- **Ziyao Cui**, Erick Jiang, Nicholas Sortisio, Haiyan Wang, Eric Chen, and Cynthia Rudin. Revisiting Broken Windows Theory. *arXiv preprint arXiv:2509.16490*, 2025.

PUBLICATIONS IN PREPARATION

- **Ziyao Cui** and Edric Tam. Learning Graph Operators for Linear Node Regression. *Expected February 2026; to be submitted to the Conference on Uncertainty in Artificial Intelligence (UAI) 2026.*

- Sahoko Ishida, Mohammed Aheed Osman, **Ziyao Cui**, Kevin Tang, and Seth Flaxman. Bayesian Small Area Estimation of Inadequate Micronutrient Intake. *Expected February 2026; to be submitted to the International Journal of Health Geographics.*

- **Ziyao Cui**, Sahoko Ishida, and Seth Flaxman. FoodSat: Small Area Estimation for Food Security Analysis with Satellite Imagery. *Expected April 2026.*

PROFESSIONAL SERVICE

Reviewer: *NeurIPS* (2025)

HONORS AND AWARDS

Honors: The Phi Beta Kappa Honor Society (Top 1% of Class), Dean's List with Distinction (4 times)

Awards: 2024 ASA Data Expo Challenge Professional Category Honorable Mention

Research Experience

Learning Graph Operators for Linear Node Regression   September 2025 – Present
Student Researcher   *Durham, NC*

- Advisor: Dr. Edric Tam, Warren Alpert Fellow at Stanford University
- Propose a two-stage framework that *learns* an interpretable graph operator from partially observed labels by minimizing a quadratic energy, decoupling operator selection from prediction.
- Introduce a positive semidefinite operator family interpolating between homophily, heterophily, and degree-driven structure, with a closed-form selection rule and clear spectral interpretation.
- Provide finite-sample guarantees under missing-completely-at-random labeling, including unbiased energy estimation, variance bounds, and sample-splitting–based regression validity.
- Demonstrate empirically that adaptively selected spectral bases improve node regression and signal compression over Laplacian and signless-Laplacian baselines, particularly in low-label and noisy-feature regimes.

How Is Uncertainty Propagated in Knowledge Distillation?   May 2025 – January 2026
Student Researcher   *Durham, NC*

- Advisor: Dr. Jian Pei, Duke University
- Lead a theoretical and empirical study of how knowledge distillation reshapes model uncertainty across linear models, neural networks, and large language models, formalizing inter-student vs. intra-student uncertainty and key uncertainty sources in the pipeline.
- Show that standard distillation suppresses within-model (intra-student) uncertainty while amplifying disagreement across students (inter-student uncertainty), leading to systematically distorted uncertainty estimates.
- Develop a bootstrap-based, data-driven framework for estimating student initialization uncertainty that complements multi-seed training and admits theoretical characterization in linear models.
- Develop variance-aware distillation methods (teacher response averaging and teacher–student variance weighting) that reduce systematic noise in LLM students by over 20% while maintaining accuracy.

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 two-parameter family of spectral embeddings that unifies Laplacian, adjacency, signless Laplacian, and related operators to interpolate between community and core–periphery structure.
- Developed spectral graph-theoretic analysis of the ILE operator family using Rayleigh quotients and quadratic forms to characterize how different hyperparameter choices encode structural patterns, and derived practical guidelines for operator and embedding-dimension selection in limited or missing-feature regimes.
- Designed and ran simulations on stochastic block models and experiments on real-world networks (Karate Club, Twitter Congress, Facebook Ego, Polblogs, Cornell, Texas, and Wisconsin), showing that ILE-based feature augmentation improves node-classification accuracy of GCN, GIN, and GraphSAGE over Laplacian baselines.

Bayesian Small Area Estimation of Inadequate Micronutrient Intake   April 2025 – Present
Student Researcher   *Oxford, United Kingdom*

- Advisor: Dr. Seth Flaxman, University of Oxford
- Develop and validate 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.
- Apply both area-level (Fay–Herriot extensions with variance smoothing) and unit-level (Beta-Binomial with spatial random effects) frameworks in a Bayesian setting, ensuring rigorous uncertainty quantification.
- Conduct validation with Rwanda's district-level representative EICV7 survey and extended methods to Senegal and Nigeria's state-level representative data, demonstrating reliable estimation at finer administrative levels where direct survey data are sparse or unavailable.
- Collaborate 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*

- Advisor: Dr. Seth Flaxman, University of Oxford
- Develop a Bayesian small area estimation framework for food security analysis, validated using Zimbabwe's ZIMVAC survey with ground-truth district-level representative data and extended to other African countries.

- Incorporate satellite imagery as auxiliary features to small area estimation models, applying computer vision models like DINO to extract high-dimensional representations that improved predictive accuracy in data-sparse regions.
- Presented at the University of Southampton to researchers from London, Oxford, Bristol, Copenhagen, and Singapore, receiving feedback on the framework's cross-modal modeling and policy implications.

On Membership Inference Attacks in Knowledge Distillation
Student Researcher

January 2025 – May 2025

Durham, NC

- Advisor: Dr. Jian Pei, Duke University
- Analyzed transferability of Membership Inference Attacks (MIAs) on 12 teacher-student LLM pairs across multiple datasets, and showed 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 architectural improvements and logit-based defenses.

Uncovering Privacy Risks in Sequential Data Releases
Student Researcher

May 2024 – October 2025

Durham, NC

- Advisor: Dr. Jian Pei, Duke University
- Demonstrated that existing privacy guarantees in individual data releases can be bypassed when sequences are jointly analyzed, revealing critical vulnerabilities in current privacy-preserving mechanisms.
- Designed and implemented 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.
- Explored 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.
- Leveraged 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: Theory and Algorithms for Machine Learning (Graduate)
Undergraduate Teaching Assistant

August 2024 – December 2024

Durham, NC

- Planned and led weekly 50-student discussions of exemplary questions to supplement course material.
- Wrote and graded homework problems (theory and coding) to assess student understanding on machine learning topics, including k-means clustering, boosting, random forest, and neural networks.
- Hosted office hours open to 200+ students, assisting with questions from homework and lectures, and answering student questions on the Ed discussion forum.

- Met weekly with the 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 – May 2025

Undergraduate Teaching Assistant *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 and machine learning in order to help train AI models.
- Evaluate and rank code generated by AI models to help cutting-edge generative 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.

COMMUNITY & LEADERSHIP

STREAM x Youth (STEM Magazine for High School Students) June 2022 - Present

Founder and Editor in Chief *Remote*

- Design structure of publication aiming to bridge the gap between high school and college STEM education.
- Provide opportunity for high schoolers to write for STEM publication and be part of STEM-lovers community.
- Coordinate a team of editorial and arts editors and oversee magazine releases, fostering a global community of high school students interested in STEM.
- Host photography and math competitions and forums to encourage international academic communications.

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)