# BRIAN LIU

briliu@mit.edu | brianliu97.github.io

## EDUCATION

**Massachusetts Institute of Technology** **September 2021 – Present**
Ph.D. Candidate in Operations Research
Advised by Professor Rahul Mazumder

**Cornell University** **May 2020**
B.S. in Operations Research | GPA: 4.1
Summa Cum Laude | Omega Rho Honor Society

## RESEARCH INTERESTS

**Topics** | Machine learning, discrete optimization, model compression, interpretability, stability
**Applications** | Clinical machine learning, pandemic modeling, healthcare operations

## SELECTED AWARDS AND HONORS

1. 2025 ISYE-MS&E-IOE Rising Star
2. 2025 American Statistical Association Statistical Computing Section Best Student Paper Award
3. 2024 INFORMS Data Mining Society Best Student Paper Award

## PUBLICATIONS

Working Papers:

1. **B. Liu** and R. Mazumder. TreePrompt: Distilling Boosted Tree Ensembles for In-Context Learning in Large Language Models, 2025.
   - Preliminary version appeared in The First Structured Knowledge for Large Language Models Workshop (KDD 2025)

Under Review:

1. **B. Liu**, R. Mazumder, and P. Radchenko. Extracting Interpretable Models from Tree Ensembles: Computational and Statistical Perspectives, 2025, Submitted to the Journal of the American Statistical Association

2. **B. Liu** and R. Mazumder. Locally Transparent Rule Sets for Explainable Machine Learning, 2025, Submitted to Operations Research

3. **B. Liu** and R. Mazumder. Randomization Can Reduce Both Bias and Variance: A Case Study in Random Forests, arxiv.org/abs/2402.12668, 2024, R&R at Journal of Machine Learning Research (JMLR)

Refereed Conferences and Journals:

1. **B. Liu** and R. Mazumder. Moss: Multi-Objective Optimization for Stable Rule Sets. In ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2025.
   - 2024 INFORMS Data Mining Society Best Student Paper Competition Winner.

2. **B. Liu** and R. Mazumder. Fast: An Optimization Framework for Fast Additive Segmentation. In ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2024.
   - 2025 American Statistical Association Statistical Computing Student Paper Competition Winner.

3. **B. Liu** and R. Mazumder. Fire: An Optimization Framework for Fast Interpretable Rule Extraction. In ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2023.

4. **B. Liu** and R. Mazumder. ForestPrune: Compact Depth-Pruned Tree Ensembles. In Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS), 2023.

5. **B. Liu**, M. Xie, and M. Udell. ControlBurn: Feature Selection by Sparse Forests. In ACM SIGKDD Conference on

Knowledge Discovery and Data Mining (KDD), 2021.

6. **B. Liu***, Y. Zhang*, S. Henderson, D. Shmoys, P. Frazier. Modeling the Risk of In-Person Instruction During the COVID-19 Pandemic. INFORMS Journal of Applied Analytics, 2024.

7. P. Frazier, J. M. Cashore, N. Duan, S. Henderson, A. Janmohamed, **B. Liu**, D. Shmoys, J. Wan, Y. Zhang. Modeling for COVID-19 College Reopening Decisions: Cornell, A Case Study. *Proceedings of the National Academy of Sciences*.

Preprints and Workshop Papers:

1. **B. Liu** and M. Udell. Impact of accuracy on model interpretations, 2020, arxiv.org/abs/2011.09903

## TEACHING

**Teaching Assistant (MIT)** **Spring 2025**
15.081 Analytics Edge (MBA)

**Teaching Assistant (MIT)** **Fall 2024**
15.072 Advanced Analytics Edge (Graduate MBAn)

**Teaching Assistant (MIT)** **Spring 2024**
15.075 Statistical Thinking and Data Analysis (Undergraduate)

**Teaching Assistant (MIT)** **Fall 2023**
15.072 Advanced Analytics Edge (Graduate MBAn)

**Teaching Assistant (MIT)** **Summer 2023**
15.067 Engineering Statistics and Data Science (Graduate LGO)

**Teaching Assistant (Cornell)** **Fall 2018 & Spring 2020**
ORIE 3300: Optimization I and ORIE 4740: Introduction to Statistical Learning.

## ADDITIONAL AWARDS

1. KDD 2025 Student Travel Grant
2. Invited to 2025 Tippie FutureBAProf Workshop

## INDUSTRY EXPERIENCE AND COLLABORATIONS

**Graduate Student Researcher** **October 2023 – Present**
SilverCloud Health | Boston, MA
Healthcare machine learning project to predict telehealth treatment outcomes for patients with anxiety and depression.

**Graduate Student Researcher** **July 2022 – Present**
Takeda | Cambridge, MA
Clinical machine learning project to improve eosinophilic esophagitis (EoE) diagnosis. Built survival models to improve the timely diagnosis of EoE and machine learning models to improve differential diagnosis.

**Graduate Student Researcher** **September 2021 – February 2022**
Hartford Healthcare | Hartford, CT
Analyzed patient and provider data to determine the impact of socio-economic factors on disparities in diabetes management.

**Data and Applied Scientist** **July 2020 – July 2021**
Microsoft (Bing Ads) | Bellevue, WA
Built and maintained machine learning models to forecast advertising account revenue and churn. Analyzed Bing demographic and interest data to optimize audience targeting for online search ads.

**Data and Applied Science Intern** **Summer 2019**
Microsoft (Devices) | Redmond, WA
Built machine learning models with >80% accuracy to forecast returns for Surface commercial customers. Designed an inventory management plan to optimize the safety stock of spare parts in the supply chain.

**Data Science Intern** **Summer 2018 & Summer 2017**
Tesla (Global Service Operations) | Fremont, CA
Built machine learning models to predict customer satisfaction with vehicle repairs. Used interpretability tools such as LIME and SHAP to identify the key drivers of customer satisfaction.

## TALKS

**MIT Sloan Health Systems Initiative Annual Workshop** **October 2024**
Interpretable Machine Learning Methods for Predicting Telemental Health Outcomes

**INFORMS Annual Meeting** **October 2024**
An Optimization Framework for Fast Additive Segmentation in Transparent ML

**ACM SIGKDD International Conference on Knowledge Discovery and Data Mining** **August 2024**
An Optimization Framework for Fast Additive Segmentation in Transparent ML

**International Symposium on Mathematical Programming** **July 2024**
An Optimization Framework for Fast Additive Segmentation in Transparent ML

**US Census Bureau Center for Statistical Research and Methodology** **July 2024**
Making Tree Ensembles Interpretable

**ACM SIGKDD International Conference on Knowledge Discovery and Data Mining** **August 2023**
Fast Interpretable Rule Extraction

**INFORMS Annual Meeting** **October 2022**
Depth-Pruning Tree Ensembles

**ACM SIGKDD International Conference on Knowledge Discovery and Data Mining** **August 2021**
Feature Selection with Sparse Forests

## OTHER RESEARCH

**Cornell Pandemic Modeling Group |** D. Shmoys, P. Frazier, S. Henderson **March 2020 – December 2022**
Modeled classroom transmission using Monte Carlo simulation to assess student and instructor risk. Analyzed the impact of university reopening, weather, and compliance to shutdowns on pandemic spread. Presented various data analyses and visualizations to Cornell senior leadership to support decisions on COVID-19 interventions. (https://datasciencecenter.cornell.edu/covid-19-modeling/)

**Cornell Pandemic Reopening Planning |** David Shmoys **March 2020 – September 2020**
Modeled student movement and behavior on campus using mobility and university internal data. Leveraged models to identify congested sidewalks and bus routes and predict potential hotspots for transmission. Tested various de-densification strategies using agent-based simulations and submitted the most effective methods to the provost. Assisted scheduling socially distanced classrooms. (https://www.orie.cornell.edu/spotlights/unsung-engineering-behind-cornells-fall-2020-schedule)

**Public Transit Route Optimization |** David Shmoys **August 2019 – March 2020**
Developed probabilistic models using farebox and vehicle location data to estimate rider demand on bus routes in Tompkins County. Used demand estimates to redesign the county's second busiest bus route to improve efficiency. Changes were implemented by Tompkins County Area Transit (TCAT) in Fall 2020.

## TECHNICAL SKILLS

Python, Julia, R, SQL, Spark, Databricks, AMPL, JuMP, Gurobi, Tableau, LaTeX

## COURSEWORK

**Machine Learning and Statistics |** Statistical Machine Learning, Statistical Learning Theory, Non-Asymptotic Statistics
**Operations Research |** Simulation Modeling and Analysis, Financial Engineering, Optimization
**Math |** Nonlinear Optimization, Real Analysis, Linear Algebra, Stochastic Processes, Probability

**Computer Science |** Database Systems, Object Oriented Programming, Software Development

## SERVICE

**ORC IAP Seminar Coordinator** **Spring 2024**

**ORC Student Seminar Coordinator** **Fall 2023**