ANGELA ZHOU

(+1) 646 918 4322 ⋄ az434@cornell.edu

EDUCATION

Cornell University *September 2016 - Present*
Fourth Year PhD Student, Overall GPA: 4.067
Department of Operations Research and Information Engineering
Undergraduate: Princeton University. Class of 2016, Operations Research and Financial Engineering. Graduated summa cum laude.

SELECTED PUBLICATIONS

Author order is alphabetical, following Operations Research convention.

Confounding-Robust Policy Evaluation in Infinite-Horizon Reinforcement Learning
With Nathan Kallus

Assessing Algorithmic Fairness with Unobserved Protected Class Using Data Combination Major Revision at Management Science.
With Nathan Kallus and Xiaojie Mao

Assessing Disparate Impacts of Personalized Interventions: Identifiability and Bounds
Proceedings of Neurips 2019.
With Nathan Kallus

Confounding Robust Policy Improvement Proceedings of NIPS 2018. Extended version accepted at Management Science.
With Nathan Kallus

Residual Unfairness in Fair Machine Learning from Prejudiced Data Proceedings of ICML 2018
With Nathan Kallus

RELEVANT COURSES

Convex Optimization, Mathematical Programming (Linear), Bayesian Machine Learning, Probability, Statistical Principles, Machine Learning and Causality for Intelligent Decision-Making, Semi/Non-Parametric Econometrics; Theoretical Machine Learning, Analysis of Algorithms, Optimal Learning, Real Analysis

EXPERIENCE

Microsoft Research New York City June 2019 - August 2019
Research Intern *New York*

· Researched optimal data collection strategies for improving inequities in machine learning model performance across groups.

My research seeks to leverage machine learning, optimization and AI to improve data-driven decision-making in important application areas. I draw on *causal inference* to study data-driven decision-making for *effective outcomes*, with strong decision-theoretic guarantees on out-of-sample causal effects. I focus on challenges that are common in practice, such as the issue of unobserved confounding that may have generated a dataset. Leveraging my background in operations research and optimization, I also draw on ideas of *credible inference* and *partial identification* from econometrics, and combining these ideas with tools in machine learning. Looking beyond guarantees for efficacy, I am also interested in ensuring guarantees for equity, with my work on fairness in machine learning, and in particular, the appropriate measurement thereof.

One line of work focuses on **learning to improve personalized decisions from observational data**. This work is related to causal policy evaluation and learning; I focus on questions of learning minimax-optimal policies in safety-critical settings such as healthcare, e-commerce, and policy. My work on "Confounding-Robust Policy Improvement" builds a case study of a parallel observational study and clinical trial to illustrate the perils of unobserved confounding. The promise of AI and machine learning allows for finer-grained adjustment for *most* confounding that generated the dataset. Our approach handles the "last-mile" of personalization. In a follow up work, "Confounding-Robust Policy Evaluation in Infinite-Horizon Reinforcement Learning", we extend the methodology for robust policy evaluation in the sequential setting.

In another line of work, I address measurement challenges for the burgeoning field of **algorithmic fairness**. In "Assessing Algorithmic Fairness with Unobserved Protected Class Using Data Combination", we develop methodology for assessing bounds on disparities in settings where the protected attribute of interest is actually not recorded in the dataset. This is common in practice. We leverage a data combination approach as well as robust optimization to provide set-valued estimates.

I am interested in building a more robust language for machine learning from observational data in order to unpack the benefits of AI *effectively* and *equitably*. Econometrics has long studied challenges of dealing with datasets generated by optimizing agents; which aptly describes the observational data landscape of contemporary health systems, tech companies, and optimizing systems.