

## EDUCATION

---

### Columbia University

Ph.D. in Computer Science, Advisor: David Blei

New York, NY

Jan 2020–Current

### Columbia University

M.S. in Computer Science

New York, NY

Sep 2018–Dec 2019

### Tulane University

B.S. in Math and Economics, B.A. in Philosophy

New Orleans, LA

Sep 2014–May 2018

– General course (2016 - 2017) at the London School of Economics

## PUBLICATIONS & PREPRINTS

---

1. A. Nazaret, **C. Shi**, D. Blei. On Misspecification in Synthetic Controls. *AISTATS 2024 Oral*
2. N. Scherrer\*, **C. Shi\***, A. Feder, and D. Blei. Evaluating the Moral Beliefs Encoded in LLMs (\*equal contribution) *Neural Information Processing Systems (NeurIPS), 2023 (Spotlight)*.
3. A. Feder\*, Y. Wald\*, **C. Shi**, S. Saria, and D. Blei. Causal-structure Driven Augmentations for Text OOD Generalization (\*equal contribution) *Neural Information Processing Systems (NeurIPS), 2023*
4. S. Casper, X. Davies, **C. Shi**, T. K. Gilbert, J.Scheurer, J.Rando, ... & D. Hadfield-Menell (2023). Open problems and fundamental limitations of reinforcement learning from human feedback. *TMLR 2023*.
5. C. Zheng\*, **C. Shi\***, K. Vafa, A. Feder, and D. Blei. An Invariant Learning Characterization of Controlled Text Generation (equal contribution) *ACL, 2023*.
6. M. Yin, **C. Shi**, Y. Wang, and D. Blei. Conformal Sensitivity Analysis for Individual Treatment Effect. *Journal of the American Statistical Association (JASA), 2022*.
7. **C. Shi**, D. Sridhar, V. Misra, and D. Blei. On the Assumptions of Synthetic Control Methods. *Conference on Artificial Intelligence and Statistics (AISTATS), 2022. (ORAL)*
8. **C. Shi**, V. Veitch, and D. Blei. 2021. Invariant Representation Learning for Treatment Effect Estimation. *Uncertainty in Artificial Intelligence (UAI), 2021. (Long talk)*
9. **C. Shi**, D. Blei, and V. Veitch. 2019. Adapting Neural Networks for the Estimation of Treatment Effects. *Neural Information Processing Systems (NeurIPS), 2019*.

### Under Review

10. M. Harrington, M. Alkon, X. He, **C. Shi**, R. Kennedy, J. Kopas, and J. Urpelainen. Anthropogenic Influences Increasingly Predict Groundwater Depletion Across India. *Environmental Research Letters (minor revision)*
11. **C. Shi\***, N. Beltran\*, A. Nazaret, A. Garriga-Alonso, A. Jesson, M. Makar, D. Blei. Circuit Checker: Hypothesis Testing for Circuits in LLMs. *ICML 2024*

## INVITED TALKS

---

- On the Assumptions and Misspecification of Synthetic Control Methods CMStatistics 2023
- Evaluating Moral Beliefs Encoded in LLMs New York Academy of Sciences – AI and Society Seminar, 2023
- Evaluating Moral Beliefs Encoded in LLMs NYU CDS Lunch Seminar, 2023
- Evaluating Moral Beliefs Encoded in LLMs Columbia workshop on text and LLMs, 2023

- Evaluating Moral Beliefs Encoded in LLMs Center for Human Compatible AI workshop, 2023
- On the Assumptions of Synthetic Control Methods. ICSA Applied Statistics Symposium, 2023
- On the Assumptions of Synthetic Control Methods. American Causal Inference Conference, 2023
- On the Assumptions of Synthetic Control Methods. Center for Causal Inference Seminar, Jan 2022
- Synthetic Controls, When do They Work and Why? Amazon Causal Inference Group, July 2021
- Nearly Invariant Causal Estimation Lyft Causal Inference Seminar, Feb 2021
- Adapting Neural Networks for Causal Estimation Oxford Machine Learning and Statistics Seminar, July 2019

## AWARDS & GRANTS

---

- **SFF Grant (\$60k)**  
Research grant to work on beneficial AI. 2020
- **Long Term Future Fund Grant (\$5k)**  
For organizing “Human-aligned AI” Official NeurIPS Social. (with Adam Gleave and Victor Veitch) 2019
- **Columbia CA Fellowship (\$15k)**  
Awarded CA fellowship for noteworthy performance as a TA 2019
- **Tulane Presidential Scholarship (\$120k)**  
Merit based scholarship going towards tuition. 2014

## PROFESSIONAL EXPERIENCE

---

**Technical Advisor** April 2022 – Now  
 FAR (<https://alignmentfund.org/>)  
 – Advise research interns on language model safety projects.

**Amazon Applied Science Intern** June 2021 – Sep 2021  
 Supervised By Dominik Janzing (received a return offer)  
 – Developed a probabilistic model and inference procedure for counterfactual estimation.

## PROFESSIONAL ACTIVITIES

---

- **NeurIPS workshop on Causal Representation Learning:** organizer, area chair 2023
- **ICML workshop on Spurious correlations, Invariance, and Stability:** organizer 2023
- **Machine Learning NYC Speaker Series:** Organizer of a city-wide speaker series and happy hours (2022 - Now)
- **Columbia ML Reading Group:** Organizer of Columbia ML reading group (2020 - 2022)
- **NeurIPS Social:** Organized Human Compatible AI social at NeurIPS 2019
- **Conference Reviewing:**
  - International Conference on Learning Representations (2021, 2022, 2023)
  - International Conference on Artificial Intelligence and Statistics (2022)
  - Neural Information Processing Systems (2020, 2021, 2023)
  - International Conference on Machine Learning (2020, 2022)