Lily H. Zhang

724-841-7941 | lily.h.zhang@nyu.edu | https://www.linkedin.com/in/lilyhzhang/ | https://github.com/lhz1029

Education

New York University — New York, NY
Doctor of Philosophy in Data Science. Advisors: Rajesh Ranganath, Kyle Cranmer. — *Aug. 2020 – May 2025*

Harvard College — Cambridge, MA
Bachelor of Arts in Statistics (primary) and Computer Science (secondary). — *Aug. 2013 – May 2017*

Honors

Graduate: DeepMind Fellow (2020), New York University Center for Data Science Graduate Fellowship (2020).
Undergraduate: Phi Beta Kappa (2017), Magna Cum Laude with High Honors (2017), John Harvard Scholar for top 5% of class (2015), Harvard College Scholar for top 10% of class (2014), Wendell Scholar nominee for "most promising and broad-ranging scholar" in the class (2014).

Publications

Conferences

- **Lily H. Zhang**, Mark Goldstein, Rajesh Ranganath. Understanding Failures in Out-of-distribution Detection with Deep Generative Models. *ICML*, 2021.
- Aahlad Puli, **Lily H. Zhang**, Eric Oermann, Rajesh Ranganath. Out-of-distribution Generalization in the Presence of Nuisance-Induced Spurious Correlations. *ICLR*, 2022.
- **Lily H. Zhang**, Veronica Tozzo, John Higgins, Rajesh Ranganath. Set Norm and Equivariant Residual Connections: Putting the Deep in Deep Sets. *ICML, 2022.*

Workshops

- **Lily H. Zhang**, Mark Goldstein, Rajesh Ranganath. Understanding Out-of-distribution Detection with Deep Generative Models. *ICLR RobustML Workshop*, 2021.
- **Lily H. Zhang** and Michael C. Hughes. Rapid Model Comparison by Amortizing Across Models. *Proceedings of the 2nd Symposium on Advances in Approximate Bayesian Inference*, 2020.

Journals

- Brittany Seymour, Rebekah Getman, Avi Saraf, **Lily H. Zhang**, Elsbeth Kalenderian. When advocacy obscures accuracy online: digital pandemics of public health misinformation through an antifluoride case study. *American Journal of Public Health.* 2015; 105(3): 517-523.

Preprints/In Submission

- Ben Townsend, Eamon Ito-Fisher, **Lily H. Zhang**, Madison May. Doc2Dict: Information Extraction as Text Generation. *Arxiv 2021.*
- **Lily H. Zhang**, Menelaos Konstantinidis, Marie-Abèle Bind, Don Rubin. Towards More Scientific Meta-Analyses.

Additional Research Experience

Undergraduate Thesis | *Advised by Jukka-Pekka Onnela* — Sep 2016 – Apr 2017
- A Comparison of Imputation Methods and Study Designs for High Frequency Survey Time Series.

HarvardX Research Assistant | *Advised by Dustin Tingley* — Jan 2016 – May 2016
- Analyzed user behavior on HarvardX and authored fakeR package.

Harvard Institute of Quantitative Social Sciences | *Advised by Gary King* — Jan 2016 – May 2016
- Explored applications of a keyword expansion algorithm that was spun into a company (Thresher).

Patents

Graphical user interface systems for generating hierarchical data extraction training dataset — Dec 2021
- US11194953B1. Co-author with colleagues from Indico Data Solutions.

Industry Experience

Student Researcher | *Google* June 2022 – Present
- Improving language models (e.g. T5) to avoid unwanted generations.

Principal Machine Learning Engineer | *Indico Data Solutions* Feb 2019 – Aug 2020
- Developed and deployed deep learning models for information extraction and document processing.

Data Scientist | *Gamalon* Apr 2018 – Feb 2019
- Built probabilistic programming-powered models for dozens of Fortune 50 companies across industries.

Data Scientist | *Jana Mobile* Jul 2017 – Apr 2018
- Co-engineered company's data science model deployment setup and built critical data pipelines.

Data Science Intern | *Premise Data* May 2016 – Aug 2016
- Developed and analyzed fraud prediction model.

Intern | *United Nations Office for Financing the Health Millennium Development Goals* Jun 2015 – Sep 2015
- Built proof-of-concept demo and helped coordinate machine learning satellite imagery extraction initiative.

Invited Talks

- Center for Data Science Graduate Student Seminar. Understanding Failures in Out-of-Distribution Detection with Deep Generative Models (2021).
- NYU AI School. Deep Generative Models (2021).
- Cornell Tech Sasha Rush and Volodymyr Kuleshov Joint Lab Meeting. Understanding Failures in Out-of-Distribution Detection with Deep Generative Models (2021).
- Fulbright Lab-to-Market Seminar. Public Health and Science Communication (2014).
- United Nations Special Envoy for Malaria Board Meeting. Lessons from Mob Malaria (2014).

Teaching

Section Leader, Machine Learning for Health Care, New York University Sep 2021 – Dec 2021
- Graduate course (CSCI-GA DS-GA 3001). Sole teaching assistant. Led and developed material for weekly sections, held office hours, and advised student projects.

Teaching Fellow, Machine Learning, Harvard University Jan 2017 – May 2017
- Undergraduate course (CS 181). Led sections and graded homework. Course evaluation of 5/5.

Academic Advising

Peer Advising Fellow, Harvard University Sep 2014 - May 2017
- Peer advisor for freshmen across backgrounds and interests. <10% selection rate.

House Advising Peer, Harvard University Sep 2015 - May 2017
- Academic peer advisor for sophomores majoring in statistics.

Other Service

- Reviewer, ICML 2022. Outstanding Reviewer (Top 10%).
- Invited Reviewer, JMLR 2022 (upcoming).
- Session Volunteer and Session Chair, ICML 2022 (upcoming).

Technical Skills

Languages: Python, SQL, R, Scala, Java, C++
Frameworks: Pytorch, Tensorflow, Pyro, Spark
Developer/HPC Tools: Git, Docker, Singularity
Libraries: Pandas, NumPy, Matplotlib, Seaborn, SciPy