

Linfeng Li

Data Scientist/Experimenter/Lecturer

✉ linfeng@umich.edu 🌐 linfeng.org 🐙 [GitHub](#) [in LinkedIn](#) ☎ +1(734)604-5530

EXPERIENCE

- Amazon** | Data Scientist and Experimenter, SEAS Feb 2023 - Present
- Built scalable data pipelines using Python, Spark, and SQL to process LLM-generated Bill of Material decomposition for over 2 million ASINs. Developed a causal model using the Heckman Selection method to link commodity price changes to vendor cost increases. Led the model's expansion, identifying \$140M in annual cost reduction opportunities across nine product categories (GLs), and delivered model predictions as a data product.
 - Employed machine learning methods (K-Means, Spectral Clustering) and developed a data pipeline to identify and extract substitution patterns from search behavior, which fed into a downstream portfolio recommendation model. Prepared the codebase for production integration and facilitated a successful handoff to the development team.
 - Engineered and administrated high-impact experiments on vendor negotiation and search behavior optimization. Leveraged AWS services to manage experimental data collection, ensuring high consistency and accuracy in data quality throughout the experiments. Developed a scalable solution for expanding the scope of the experiment, and recruited over 50,000 participants in two weeks.
- University of Michigan** | Postdoc Researcher Nov 2022 - Present
- Coordinate a large-scale field experiment with Digital Green, drawing on experience with Python, oTree, and JavaScript to establish a lab-in-the-field study focused on the adoption of an LLM-assisted chatbot, Farmer.Chat, in India.
 - Oversee lab operations at the Behavioral Laboratory, including participant recruitment, experiment setup, and providing critical feedback on experiment designs.
- University of Michigan** | Lecturer I (for the Master of Applied Data Science program) May 2022 - Present
- In Causal Inference (SIADS 630), cover topics on controlled regression, matching, IV, RD and DID.
 - In Data Science for Social Good (SIADS 688), cover applications of causal inference and behavioral economics theories to enhance pro-social behaviors across diverse domains (e.g. energy conservation and microfinance).
- Amazon** | Economist Intern, CoreAI Jul 2022 - Oct 2022
- Translated a theoretical model of seller/vendor contract selection into an experimental interface, implementing both front-end and back-end components in oTree to create a controlled decision environment. Leveraged SQL to pull and calibrate with real business data. Administered the experiment end-to-end, providing insights on optimal contract design.
- University of Michigan** | Research Associate, Empirical Legal Studies Center Aug 2014 - Aug 2015

SKILLS

Methods | Causal Inference (Experiments, DML, DID, RDD, Synthetic Control), Machine Learning (K-Means and Spectral Clustering, PCA, Linear/Logistic/Lasso/Ridge Regression, Random Forest, Boosting/Bagging, Ensemble Methods), Social Network Analysis (community detection), Survey Design

Packages | Pandas, PyTorch, PySpark, Numpy, SciPy, sklearn, linearmodels, Stargazer

Languages | Python, SQL, Stata, \LaTeX | **Technologies** | Linux, Vim, Tmux, oTree, ORSEE, ArcGIS, Git

EDUCATION

- University of Michigan** | Ph.D. in Information 2015 - 2022
Thesis: Essays on Field Experiments in Behavioral Economics
- University of Michigan** | Bachelor of Science in Pure Math, Economics (with Honors) 2012 - 2014
with minor in Complex System
- Central University of Finance and Economics, Beijing, China** | Bachelor of Science in Mathematical Economics and Mathematical Finance (transferred) 2010 - 2012

PUBLICATIONS

- M.-J. Fong, J. Li, **L. Li**, D. Bergemann, Y. Chen. Vendor Negotiation Experiment and Training Tool. 11th Annual Consumer Science Summit, July 2024 (Amazon Internal).
- **Li**, Dillahun, and Rosenblat. Does Driving as a Form of "Gig Work" Mitigate Low-Skilled Job Seekers' Negative Long-Term Unemployment Effects? Proc. ACM Hum.-Comput. Interact. 3, CSCW 2019
- Dillahun, Kameswaran, **Li** and Rosenblat. Uncovering the Values and Constraints of Real-time Ridesharing for Low-resource Populations. CHI 2017, ACM, New York, NY, USA

WORKING PAPERS

- Motivating metadata contributions for data reuse and reproducibility, a field experiment (with Yan Chen, Margaret Levenstein and Lars Vilhuber, draft available)
- Increasing and retaining women's participation in Wikipedia, a field experiment (with Yan Chen, David Miller and Muriel Niederle, draft available)