Shuoqi Sun

Email — Github — LinkedIn — Personal website

Education

Carnegie Mellon University PhD in Economics (STEM, Dec.2025)
Columbia University in the city of New York MA in Quantitative Methods in the Social Sciences
Peking University BA in Philosophy(Logic Track) & Economics

Analytic Experience

ByteDance *Data Science Intern, E-Commerce Governance and Experience*
May-Sep. 2025

- **Developed a multi-task ML model** for merchant lifetime prediction and uplift estimation of experience costs using causal machine learning methods, trained on millions of active sellers and multi-terabyte transaction-level data. Achieved forecast MAPE 10% lower than benchmark.
- **Built automated pipelines** for daily data ingestion, retraining, and dashboard integration, supplying key coefficients to the ROI framework to quantify long-term merchant outcomes and evaluate governance strategies.
- **Analyzed marketplace experiments to quantify ecosystem effects and set guardrails:** Combined experiment insights with model-derived coefficients to uncover cross-side dynamics between buyers and sellers, and estimate an upper bound of governance intensity beyond which long-term GMV loss is likely, providing quantitative guardrails for sustainable strategy design.
- Researched cutting-edge identification strategies—including variance reduction techniques, de-biased estimators, and scalable panel data methods—to improve credibility in causal inference with observational data.

Carnegie Mellon University, Tepper School of Business *Doctoral Researcher*
2020-2025

- Leverage economic theory(game theory, mechanism design) and statistical data analysis to understand the equilibrium effects of personalization and privacy regulations on marketplace ecosystem and optimize business strategies.

Selected Data Science Projects

Autonomous causal measurement agent

- Designed and implemented an agentic system that autonomously formulates causal estimands, selects identification strategies (e.g., difference-in-differences), and orchestrates multi-step measurement workflows using LLMs.
- The agent performs diagnostic checks (e.g., parallel trends), evaluates the credibility of estimated effects, and iteratively refines measurement decisions or recommends experimentation when observational evidence is insufficient.

Two-Stage News Recommendation System (Recall + Ranking)

- Designed a two-stage news recommendation system (recall + ranking) on 300K users and 3M interactions, trading-off between retrieval efficiency and ranking quality under real-world constraints.
- Built hybrid recall mechanisms using deep user/content embeddings with approximate nearest neighbor search, and engineered behavioral and content-based features to support downstream ranking models and offline evaluation on noisy, implicit-feedback data.

Does recommendation density accelerate or delay conversion?(Mercari marketplace,WIP)

- Constructing embedding-based item similarity and a "recommendation density" measure using text metadata to quantify each listing's exposure quality within the Mercari ecosystem.
- Modeling two opposing forces—faster matching vs. increased comparison shopping—to understand how recommendation quality affects item-level time-to-sale and marketplace liquidity.
- Using heterogeneity in price elasticity to identify the causal impact of recommendation density.

Technical Skills

Machine Learning: supervised and unsupervised learning, generalized linear models, tree-based methods (LightGBM, XGBoost), neural networks, dimensionality reduction (PCA), time-series modeling
Causal Inference: A/B testing, causal ML(DML, meta learners, DR learner, Causal forest), DiD, PSM, IV
Data, Engineering & Agentic Systems: Python (Pandas, Scikit-learn, PyTorch, PySpark), SQL, Spark, Hive, Git, Tableau; scalable data pipelines; LLM-based agentic workflows (LangGraph), programmatic LLM APIs