Yingqi Gao

Machine Learning Researcher | PhD Candidate in Statistics, UCLA

yingqi.gao.stats@gmail.com | 858-666-7501
yingqi-gao.github.io | linkedin.com/in/yingqi-gao

PhD-trained **data scientist** and **researcher** designing **incentive-aligned**, **profit-aware learning systems** grounded in **economic theory** and **mechanism design**, leveraging **machine learning**, **reinforcement learning**, **optimization**, and **causal inference** to advance **data-centric decision making** under **uncertainty**.

Work Experience

PhD Researcher – Machine Learning & Mechanism Design   Feb 2023 - Present
University of California, Los Angeles (UCLA) | Los Angeles, CA

- Designed an auction-to-posted-price mechanism for **ML data markets**, integrating **SciPy** optimization and **R**'s nonparametric estimation, achieving no-regret, data-efficient pricing with **2× faster convergence**.
- First-authored *Learn then Decide: A Learning Approach for Designing Data Marketplaces* (**major revision at JASA**, Theory & Methods), presenting the mechanism and pricing frameworks, and now exploring an **LLM-based market** for **time-series privacy**, calibrated for incentive alignment.
- Built a **PyTorch** framework for prediction-powered **uncertainty quantification** in **high-dimensional causal mediation**, benchmarked against **XGBoost** and **DoubleML**, and scaled to **Google Cloud**.

Machine Learning Researcher   Sep 2022 - Dec 2022
University of California, Los Angeles (UCLA) | Los Angeles, CA

- Engineered a **PyTorch**-based framework for object-centric **reinforcement learning**, extending OP3 with **variational encoding** and **attention**-based decomposition to enable **zero-shot generalization** and interpretable multi-object control across dynamic environments.

Graduate Researcher   Jun 2020 - Jun 2021
Columbia University | New York, NY

- Optimized a **Bayesian hierarchical sparse VAR model** for multi-subject, multi-session **fMRI** analysis using **R** and **HPC**, cutting **runtime by 80% (15→3 days)** through efficient **Gibbs sampling** and enhancing stability through sensitivity analysis.

Core Skills

Python (PyTorch, TensorFlow, SciPy, scikit-learn, XGBoost), R, C, C++, HPC, Google Cloud, Git, SQL.

Education

University of California, Los Angeles (UCLA)   Sep 2021 - Dec 2026
PhD in Statistics & Data Science

Columbia University   Sep 2019 - Dec 2020
MA in Statistics

University of California, San Diego (UCSD)   Sep 2015 - Jun 2019
BS in Probability & Statistics; **BS** in Management Science, Graduated with High Honors