

Jiayou (Rick) Liang

+86-13829921415 | jiayouliang@uchicago.edu | jiayouliang@hotmail.com | rickliang-jy.github.io | GitHub: [rickliang-JY](https://github.com/rickliang-JY)

EDUCATION

University of Chicago Chicago, IL
M.S. in Applied Data Science | GPA: 4.0/4.0 | Oct 2024 - Dec 2025

Hong Kong Baptist University Zhuhai, China
B.S. in Financial Mathematics, First Class Honours | GPA: 3.77/4.0 | Sep 2020 – Jun 2024

University of Oxford (Online) Aug 2022
Summer Course: Artificial Intelligence and Machine Learning (Grade: A)

SKILLS

- **LLM & AI Algorithms:** Fine-tuning, Transformer architectures, optimization algorithms, prompt engineering, inference optimization, vector databases (FAISS)
- **Machine Learning:** Deep learning, reinforcement learning, supervised / contrastive / self-supervised learning, computer vision, combinatorial optimization, integer programming
- **AI Agent Development:** LangChain, LangGraph, AutoGen, RAG (Retrieval-Augmented Generation), Azure Cognitive Search, multi-agent system design
- **Programming:** Python (PyTorch, scikit-learn, NumPy, Pandas), C++, SQL (PostgreSQL, MySQL), R, Bash
- **Tools & Platforms:** Git, Docker, Google Cloud, Azure, Tableau, LaTeX, Claude Code

RESEARCH EXPERIENCE

Accelerating Large-Scale Adaptive Optimization via Reparameterization Apr 2025 – Present
University of Chicago | Advisor: Prof. Rebecca Willett

- **Algorithm design:** Led the design of Auxiliary-Variable and Reduced EGOP algorithms to bypass the storage/compute bottlenecks of full-matrix EGOP reparameterization; theoretically tightened Adagrad/Adam convergence bounds by a factor of $1/d$.
- **Low-rank + orthogonal complement:** Used randomized SVD to extract the top- r eigenvectors of the EGOP matrix; introduced an orthogonal-complement auxiliary variable enabling lossless full-space optimization while storing only a small basis.
- **Empirical results:** Implemented the module end-to-end and benchmarked on Fashion-MNIST, EMNIST, CIFAR-10; on a 78,400-dim layer, retaining $<1.3\%$ of components ($r=1000$) — and even $r=50$ — captured the loss landscape's dominant geometry and accelerated convergence.

Neural Assortment Optimization (NAO) Apr 2025 – Present
University of Texas at Austin | Advisor: Prof. Rui Gao

- **Full-stack development:** Built the NAO codebase from scratch, implementing an "extend-particle search-round" pipeline: Lovász tight extension lifts the discrete objective onto $[0,1]^n$ with exact optimal value, multi-particle projected noisy SGD with entropic-risk pooling explores the non-convex landscape, and chain rounding recovers a discrete assortment with no rounding-error term. Supports BAM, NL, and MMNL choice models.
- **Scaling & theory:** Designed a rolling-window construction for capacity-constrained settings that maintains feasibility while preserving informative subgradients. Pushed tractable size to $n=10,000$ items at near-0% optimality gap on public benchmarks.
- **Beating baselines:** Outperformed SOTA heuristics (e.g., ADXOpt) and neural baselines (NN/NNpp) on MMNL stress tests; at 10k+ scale, maintained fastest runtime and lowest error while exact solvers like Gurobi (conic integer formulation) hit OOM or exponential blow-up.

Preference Alignment of a Pathology Vision Foundation Model — Capstone (Honorable Mention) Apr 2025 – Dec 2025
University of Chicago & Medical School | Advisors: Prof. Utku Pamuksuz, Dr. Samir Atiya

- **Vision-DPO architecture:** First to transfer Direct Preference Optimization from LLM alignment to computational pathology. Designed a pseudo-probability framework (cosine similarity \rightarrow temperature scaling $\gamma=10 \rightarrow$ sigmoid) converting TITAN encoder geometry into DPO preference likelihoods; LoRA ($r=16$) trains only 2.7% of base parameters with frozen reference model for implicit KL regularization.
- **Self-supervised preference data:** Designed a label-free preference-pair pipeline: baseline classifier flags misclassified samples; UMAP + DBSCAN over correct clusters yields "preferred" embeddings vs. misclassified "rejected" targets. End-to-end: preprocessing \rightarrow preference generation \rightarrow LoRA fine-tuning \rightarrow SCL \rightarrow classification.
- **Results:** Lifted PR-status AUROC from 79.75% \rightarrow 81.92% and Cohen's κ from 45.40% \rightarrow 54.43% over raw-TITAN baselines; awarded Honorable Mention at the UChicago Applied Data Science Capstone Showcase.

SELECTED PUBLICATIONS

- Liu, R., **Liang, J.**, Chen, H., Hu, Y. (2025). "Analyst Reports and Stock Performance: Evidence from the Chinese Market." Asia-Pacific Financial Markets. — NLP, BERT, sentiment analysis
- DePavia, A., Cruzado, J., **Liang, J.**, Charisopoulos, V., Willett, R. "Faster Adaptive Optimization via Expected Gradient Outer Product Reparameterization." (Submitted to SIMODS.) — Adam/Adagrad, low-rank preconditioning
- Atiya, S., **Liang, J.**, et al. "Validating Two Pathological Foundation Models for Breast Biomarker Detection." (In preparation.)

PROJECT EXPERIENCE

AI-Assisted Clinical Data Navigation Platform (LLM + Agentic AI)

St. Jude Children's Research Hospital

- **Agentic NL-to-SQL interface:** Designed and built an LLM agent chat interface enabling researchers to explore the Childhood Cancer Survivor Study (15,000+ de-identified variables across survivor and sibling cohorts) via natural-language questions, auto-translated to SQL with instant summary statistics and visualizations — dramatically lowering the barrier to pre-proposal data exploration.
- **RAG + multi-agent retrieval:** Built a Retrieval-Augmented Generation pipeline integrating Azure Cognitive Search with a LangChain / LangGraph multi-agent framework for intelligent variable discovery, codebook lookup, and document retrieval, replacing manual hunting through legacy data dictionaries.
- **Production deployment:** Delivered the end-to-end stack — secure PostgreSQL backend with role-based access control, LLM API integration, prompt orchestration, and automated data-processing flows — supporting both cohorts in a HIPAA-conscious clinical research environment.

[HikerScrolls](#) — Outdoor Hiking Scrollytelling Plugin

Obsidian Community Plugin (Author)

- **Heterogeneous data fusion:** Independently designed and shipped a published Obsidian community plugin parsing and fusing GPX track data, EXIF photo metadata, and 10+ map tile sources (Stadia, OpenStreetMap, etc.); rendered three visualization templates: scrollytelling narrative, photo scrapbook, and hand-drawn map.
- **Multimodal LLM integration:** Integrated the Gemini LLM for AI trip-summary generation and AI-driven layout optimization; built a "Souvenir Store" module driven by multimodal prompt engineering to generate 5 categories of travel souvenirs (postcards, fridge magnets, etc.) from trip photos and metadata.
- **Full product surface:** Architected the Journal Wizard, elevation-profile renderer, global Atlas map view, timeline sidebar, and bilingual layer; published to the official Obsidian community plugin store.

[Cross-Asset Order Flow Imbalance \(OFI\) Analysis](#)

Quant Research Project

- **HFT feature engineering:** Engineered 5-level Order Flow Imbalance features from high-frequency limit-order-book data on 5 cross-sector Nasdaq 100 stocks (AAPL, AMGN, TSLA, JPM, XOM); used PCA to aggregate the multi-level signals into a unified metric, raising the average self-impact regression R^2 from 0.54 to 0.65.
- **Cross-asset impact modeling:** Built four contemporaneous regression families (PI/PII self-impact, CI/CII cross-impact) to quantify how a single stock's OFI propagates to other names' price movements; the cross-asset integrated CII model achieved best-in-class performance across all stocks (avg $R^2 = 0.71$, max 0.75 on AMGN), empirically validating order-flow information transmission in multi-asset pricing.

INTERNSHIP EXPERIENCE

Exion Asia (Huizhou) Co., Ltd. — Marketing & Sales Analyst Intern

Jun 2023 – Aug 2023

- Decomposed sales/revenue trends by region, industry, and product line; identified seasonal and promotional drivers for executive weekly and monthly reports.
- Built multi-dimensional Tableau dashboards (KPI tracking, sales funnel, geographic distribution), automating the weekly/monthly reporting workflow.

Bank of Communications, Huizhou Branch — Customer Service Manager Intern

Jun 2022 – Jul 2022

- Rotated through teller and lobby positions serving 200+ customers with zero complaints; supported wealth managers in client needs analysis and product recommendation.

AWARDS & HONORS

- Honorable Mention — University of Chicago Applied Data Science Capstone Showcase (2025)
- First-Class Scholarship — BNU-HKBU United International College (2021, 2022, 2023); President's Honor Roll
- Third Prize — Guangdong Division, China Undergraduate Mathematical Contest in Modeling (2022)