

Zijun Liu

San Diego, CA · liuzijun6688@gmail.com · 530-220-8681 · [Github](#) · [Portfolio website](#)

Summary

UCSD Astrophysics & Data Science student focused on systematic quantitative research and AI-driven signal generation. I architect production-grade ML pipelines and low-latency C++ systems, applying rigorous strategy falsification, regime-switching logic, and statistical modeling to distinguish true predictive edge from overfitting noise.

Education

University of California, San Diego - B.S. Astrophysics, Data Science minor, *Sep 2025 ~ June 2027*, **GPA 3.98 (institutional)**

University of California, Davis - B.S. Physics (Astrophysics Emphasis) & Data Science, *Sep. 2023 ~ June 2025*, **GPA 3.84**

Skills/Tools

- **Languages:** Python, C++, R, SQL, Bash/Shell
- **Quantitative Methods:** Strategy Falsification, Vectorized Backtesting, Market Microstructure, Risk Analysis,
- **Machine Learning & AI:** PyTorch, TensorFlow, Hugging Face (TRL, Transformers), Model Quantization (4-bit), PEFT/LoRA, XGBoost, LightGBM, Temporal Convolutional Network (TCN), Scikit-learn, Optuna
- **Data Engineering:** DuckDB, Polars, Pandas, PostgreSQL, Multiprocessing, ETL Architecture
- **Cloud & MLOps:** AWS (Lambda, S3), Git/GitHub CI/CD, Docker, Linux Environment

Experiences

Rothenberg Wealth Strategies — AI/ML Quantitative Research Intern (*Jan 2026 – Present*)

- **Engineered a ETL pipeline** utilizing **DuckDB** and **Polars** for rapid pre-filtering and **Pandas** for granular manipulation, processing 12 years of daily market data (**tens of millions of rows**) while resolving GPU I/O bottlenecks via asynchronous **PyTorch DataLoaders**.
- **Architected an 8-layer Temporal Convolutional Network (TCN)** for 20-day forward return probabilistic quantile forecasting (P10/P50/P90); stabilized training via **gradient clipping** and **pruned 200+ features** using LightGBM mutual information to mitigate overfitting.
- **Conducted falsification on 40 LLM-generated alpha factors** sourced from recent academic literature; implemented out-of-sample vectorized backtesting and prevented deployment of spurious models by proving all 40 signals yielded negative returns after introducing transaction costs.

Himiway Intelligent Technology USA - Data Science intern (*Sep 2025 – Dec 2025*)

- **Architected a scalable SQL/Python ETL pipeline** for e-bike sales forecasting; refactored experimental Jupyter notebooks into a production-ready, modular codebase (argparse, robust logging, Makefile orchestration) to streamline continuous data ingestion.
- **Developed and deployed an end-to-end XGBoost model** to optimize regional store locations and inventory allocation, successfully accelerating the executive decision-making cycle by **75% (from 4 weeks to 1 week)**.
- **Designed and evaluated a months-long strategic A/B test** on in-store customer experiences (fat-tire vs. thin-tire); leveraged **Tableau** dashboards to analyze key metrics including Conversion Rate and Revenue per Visitor, driving data-informed retail operations.

Projects

KataGo × LLM - Explainable Go AI | *Project Lead - 2025 ~ Present*

- Engineered a **regime-switching** RL (GRPO) pipeline, adaptively overweighting rank-based signals in high-uncertainty states (win-rate ~0.5) and policy priors in deterministic regimes to optimize risk-adjusted decision making.
- **Mitigated reward hacking** by implementing strict legality gates and downsampling imbalanced, low-information data (lopsided positions), elevating the model's performance from baseline to a ~7k rating over a 31-hour run.
- **Architected a zero-network-overhead C++ proxy** for real-time state evaluation and quantized the 8B model to 4-bit GGUF, achieving ultra-low latency inference (0.52s TTFT, 42.5 tok/sec) in resource-constrained environments.

ASTR199 - Stellar Parameter Prediction via Deep Learning (with Prof. Theissen, UCSD) | *Machine Learning Researcher - 2026*

- **Engineered orthogonal luminosity features** to break prediction degeneracy after identified PCA's info loss on highly collinear inputs; implemented 1-sigma outlier clipping across F/G/K/M cohorts to robustly filter background noise.
- **Architected a two-stage DL pipeline** for multi-target prediction, resolving multi-task loss collapse. Strictly enforced train/val/test isolation between pipeline stages to **prevent data leakage**.
- Conducted residual analysis to **define model boundaries**, diagnosing specific prediction failures (K-dwarfs) as a consequence of narrow target variance and long-tail sample imbalance rather than model capacity constraints.

Breaking Barriers Hackathon (AWS × Deloitte × AT&T) – **Finalist (Top 8/32)** - 2025

- Architected a real-time anomaly detection system for simulated tail-risk events, implementing a **0.5-1s micro-batching pipeline** to efficiently process LLM-generated synthetic data streams.
- Engineered **lagged spatio-temporal features** via client-side sliding windows for stateless backend processing; trained an **XGBoost classifier**, tuning thresholds via **PR AUC** to minimize False Negatives on imbalanced data.
- Deployed ultra-low latency model inference via **AWS Lambda (<100ms)**, streaming anomaly triggers to an S3-hosted **Amazon Location Service** interactive dashboard for real-time visualization.

PHY199 - Galaxy Spectra PCA Pipeline (with Prof. Wittman, UCD) | *Researcher - 2025*

- **Developed an automated Python/R preprocessing pipeline** to clean 2,300+ high-dimensional spectra, implementing custom unsharp masking to resolve artifacts and anomalous flux spikes.
- **Designed an unsupervised Machine Learning framework** using **scikit-learn (PCA)** for automated outlier filtering, eliminating massive manual data inspection and extracting key features (PC1-PC3).
- **Packaged the analytical workflow into a reproducible codebase**; generated automated visualizations in R to translate complex high-dimensional signals into actionable insights.

Leadership Experience

Go Club at UC Davis - President & Co-Founder (Sep 2024 - June 2025)

- Grew to **30+ members**; coordinated with ACGA and universities to run inter-school competitions and community events.
- Led cross-cultural, mixed-skill communications; organized weekly programs and external collaborations.