Shizhe "Andy" Zhang

shizhe_zhang@berkeley.edu · (+1) 510-646-3962 · Website
Berkeley, California, USA

Education

University of California, Berkeley | GPA: 4.0/4.0 08.2025 - 12.2026

M.A. in Statistics

Advanced topics in Neural Networks, NLP, Statistical Computing, Statistical Learning Theory, Optimization Models in Engineering

Zhejiang University | GPA: 3.96/4.0 08.2021 - 06.2025

B.S. in Economics (*Advanced Class*, Selected from top 10%)

Game Theory, Incentive Theory, Econometrics, Statistical Software, Time Series, Quantitative Investment, Big Data Analysis
Minor (Mathematics): Real Analysis, Stochastic Process, Numerical Analysis, PDE, Statistical Learning, Data Structure
Honor: First Prize in National College Mathematics Competition, certificate of the chu kochen honors program (top 1.3%), First Prize Scholarship for Outstanding Econ Students (top10%), Outstanding Graduates, College-level Scholarship

Skills

- **Programming:** Python, R, SQL, Matlab, SQL (MySQL, PostgreSQL), Shell Scripting (Bash, Zsh)
- **Tools & Software:** Git, GitHub Actions, Vim, Unix, Web Scraping, Cloud Computing (AWS, GCP, PSC), Latex
- **Technical Skills:** Feature Engineering (Pandas, NumPy), Data Visualization, Machine Learning (Scikit-learn), Deep Learning (PyTorch, HuggingFace), Causal Inference, Optimization (CVXPY), Model Evaluation & Validation

Internships

Guotai Haitong Securities, Quantitative Researcher 06.2025 - 08.2025

- Generated event-driven strategies from the A-share Billboard Stocks (*Dragon-Tiger List*).
- Built predictive features from raw trade data, including net inflow strength, institutional concentration, and order imbalance.
- Trained a LightGBM to predict short-term price movements. Top-quintile portfolio with a monthly excess return of 2.3% over the market benchmark.
- Constructed a capital-activity factor by modeling synergistic trading behavior using brokerage network analysis. Deployed the factor in a long-short strategy (CSI 300 vs. CSI 2000), boosting signal precision by 14% over heuristic benchmarks.

SDIC Securities, Research Assistant 06.2024 - 08.2024

- Engineered an automated ETL using Python and SQL to process electricity spot/futures data for downstream modeling. Implemented the automated calculation of market indices (e.g. Price Spread, Volatility) as reliable sources.
- Developed a proxy for CUR(Capacity Utilization Ratio) feature with an XGBoost model on alternative data(e.g., logistics data, supply chain metrics); the feature ranked among the top 5% impactful predictors in an earnings forecasting model.

Projects

Rag-Driven Company Search System (Industry Partner: Pyramyd) 10.2025 - 12.2025

- Built RAG-based discovery system over 100K+ reviews, integrating keyword search, embeddings, and FAISS retriever.
- Conducted sensitivity and counterfactual ablation analyses to validate robustness, demonstrating 85–90% top-k ranking stability and 20–35% Recall@K degradation when key components were removed.
- Implemented a modular experimentation pipeline spanning data parsing, chunking, indexing, caching, grounding, and evaluation, enabling systematic ablations across 10+ retrieval and reranking variants while avoiding full reindexing for 80% of experiments.
- Delivered strong system performance using hybrid candidate fusion and metadata reranking, improving retrieval relevance by 30-50% (Recall@K / NDCG@10), achieving real-time (<1s) retrieval, and reducing average LLM prompt size by 40%.

Robust Portfolio Optimization of CNE5 Factors 11.2024 - 12.2024

- Standardized 100+ raw style/industry factors, applied sparse PCA factor selection to derive interpretable and stable factor structures, and estimated factor covariance matrices under high dimensionality.
- Optimized portfolios using mean-variance and CVaR minimization under sector-neutrality and turnover constraints.
- Evaluated robustness via 10,000+ Monte Carlo stress scenarios and block bootstrap resampling,
- Benchmarked Barra-style linear risk models against ML-enhanced covariance estimation, including Ledoit-Wolf shrinkage and Graphical Lasso for high-dimensional stability, improving Sharpe ratio by 11% and reducing max drawdown by 6%.

Capital Market, Rational Inattention and Internet Search 12.2023 - 02.2024

- Developed a distributed, asynchronous web-scraping pipeline using Python (Scrapy, Selenium) to construct a large-scale panel dataset of investor attention, capturing 5000+ daily observations of Baidu Search Index since 2013.
- Engineered a suite of over 10 attention factors. Orthogonalized these factors against the Fama-French 5-factor model using Gram-Schmidt decomposition, boosting the out-of-sample R^2 by 12.4% for 1-week forward returns.
- Quantified the causal impact of attention spikes on asset prices using DID and IV methodologies, demonstrating that a one-standard-deviation increase in search intensity yielded a 0.38% abnormal 3-day return.
- Benchmarked linear factor models against deep learning baselines for return prediction, including LSTM and GRU models, which achieved a 15% lower MSE particularly during high-volatility regimes identified by a GARCH filter.

Leadership

President, Billiards Association, Zhejiang University 06.2022 - 06.2025

- Responsible for club promotion, expanded membership to 200+; organized tournaments and secured sponsorships.