

# Zhengze Zhang

✉ [zhangzhengze2018@gmail.com](mailto:zhangzhengze2018@gmail.com) | ⚡ [github.com/Garthzzz](https://github.com/Garthzzz) | 🌐 [garthzzz.github.io](https://garthzzz.github.io)

## Education

<b>Columbia University</b>	New York, USA	Sep 2024 – Present
<i>M.A. in Statistics (Machine Learning Track)</i>		
<ul style="list-style-type: none"><li>Courses: Statistical Machine Learning, Natural Language Processing, Advanced Data Science, Financial Statistics, Linear Regression, Statistical Inference, Statistical Computing and Data Science.</li></ul>		
<b>University of California, Santa Barbara</b>	Santa Barbara, USA	Sep 2018 – Jun 2023
<i>B.S. in Mathematics &amp; B.A. in Physics</i>		
<ul style="list-style-type: none"><li>Mathematics: Real Analysis, Complex Analysis, Differential Geometry, Regression Analysis, Bayesian Analysis, Estimation Theory.</li><li>Physics: Quantum Mechanics, Statistical Mechanics, Electromagnetism, Astrophysics, Theoretical Methods, Scientific Programming.</li></ul>		

## Research Experience

<b>Columbia University</b>	New York, USA	
<b>Constrained Symbolic Regression for Financial Return Prediction</b>		
<i>Advisors: Prof. Tian Zheng, Dr. Kara Lamb, Dr. Mikhail Smirnov</i>		
		Oct 2025 – Present
<ul style="list-style-type: none"><li>Developed Constrained Symbolic Regression (CSR) framework integrating Campbell &amp; Thompson (2008) sign constraints with task-aware wavelet denoising and stability selection for equity premium prediction using Goyal-Welch macroeconomic predictors.</li><li>Achieved interpretable closed-form equations with economically meaningful coefficients; validated cross-domain generalizability on Treasury yield curve modeling (<math>R^2 &gt; 95\%</math>), Fama-French factor timing, and HAR volatility forecasting.</li><li>Conducted rigorous validation via permutation tests (Sharpe <math>p=0.012</math>), placebo constraints, sub-period stability analysis, and bootstrap confidence intervals; identified signal-to-noise ratio as the critical success factor across financial domains.</li></ul>		
<b>Cloud Microphysics Emulation: Next-Generation Symbolic Regression Algorithms</b>		Aug 2025 – Present
<i>Advisor: Dr. Kara Lamb, LEAP Center, Columbia University</i>		
<ul style="list-style-type: none"><li>Designed three-stage symbolic regression framework integrating E-WSINDy (Ensemble Weak-form SINDy), LaSR (Language-Augmented SR), and AI Feynman 2.0, achieving 50-1000x noise robustness improvement by transferring derivatives from noisy data to smooth test functions via weak-form integral formulations.</li><li>Applied Hessian-based decomposition to 9-dimensional warm rain microphysics data, proving additive structure in autoconversion processes; integrated LARS variable selection reducing search space from 9 to 4-6 variables with 4-6x computational speedup.</li><li>Implemented Buckingham <math>\pi</math> theorem for dimensional analysis, reducing feature space to 5-6 dimensionless parameters ensuring physical consistency; developed bootstrap ensemble methods for uncertainty quantification with physics-informed conservation constraints.</li></ul>		
<b>Weakly Supervised Tree Species Classification via Algorithm Optimization</b>		Sep 2025 – Present
<i>Advisor: Prof. Tian Zheng, TZstats Convergence Lab</i>		
<ul style="list-style-type: none"><li>Migrated full APL pipeline from TensorFlow to PyTorch with GPU acceleration, processing 14,400 image patches for tropical forest tree species identification using weakly supervised learning with imprecise point labels.</li><li>Proposed algorithmic improvements: replaced K-means with graph-based clustering (kNN + Leiden) and density-based clustering (HDBSCAN) to capture irregular embedding manifolds; designed sigmoid-weighted loss using continuous relevance scores for robustness under noisy annotations.</li><li>Analyzed APL's limitations on general image domains and proposed adaptation strategies including multi-scale backbones, transition from density regression to instance detection, and MIL/CAM-based weak supervision.</li></ul>		
<b>Treatment Discontinuation Modeling in CATIE Antipsychotic Trial</b>		May 2025 – Present
<i>Advisor: Prof. Kiyohito Iigaya, Iigaya Lab</i>		
<ul style="list-style-type: none"><li>Established reproducible EDA and preprocessing pipelines for CATIE clinical data, integrating demographics, PANSS scores, adverse events, and neurocognitive assessments across 1,400+ schizophrenia patients.</li><li>Applied PCA and factor analysis to identify latent symptom dimensions predicting Phase I discontinuation due to inefficacy or intolerance; modeled discontinuation risk via regression linking symptom factors and treatment variables. Manuscript in preparation.</li></ul>		
<b>Lineage Heterogeneity and Founder-State Modeling in Monoclonal Gastruloids</b>		May 2025 – Present
<i>Advisor: Prof. Bianca Dumitrascu, Morpho Lab</i>		
<ul style="list-style-type: none"><li>Studied DNA Typewriter lineage-tracing methods (Regalado et al., 2025) to inform founder-state modeling in developmental biology.</li><li>Processed gastruloid scRNA-seq data and lineage trees using PCA/UMAP; computed pairwise lineage distance matrices and quantified inter-gastruloid heterogeneity through lineage distance–fate similarity relationships.</li></ul>		
<b>ALFIE and VVIQ Mental Imagery Profile Analysis</b>		Jan 2025 – Present
<i>Advisor: Dr. Alfredo Spagna, The Living Lab</i>		
<ul style="list-style-type: none"><li>Integrated VVIQ and ALFIE task scores from 100 participants; employed ensemble clustering and co-clustering consensus to identify three robust imagery phenotypes, validated by silhouette scores and bootstrap-adjusted Rand index. Manuscript in preparation.</li></ul>		
<b>Sentiment Dynamics in COVID-19 Tweets: Classical vs. Transformer Models</b>		Sep 2024 – Dec 2024

*Advisor: Prof. Patrick Houlihan*

- Analyzed 364,802 tweets comparing lexicon-based (VADER, LIWC) and transformer-based models (RoBERTa); used OLS and PCA to link sentiment scores with user features, demonstrating that contextual embeddings better capture subtle sentiment dynamics.

### **University of California, Santa Barbara**

Santa Barbara, USA

#### **Photometric Analysis and Classification of Supernova AT2023hpb**

Mar 2023 – Jun 2023

*Advisor: Prof. Phillip Lubin, UCSB Experimental Cosmology Group*

- Conducted deep observation of AT2023hpb using Las Cumbres Observatory global telescope network; processed raw images via AstroArt 8 and Atlas; analyzed signal-to-noise ratios and plotted light curves in Python to infer Type II classification.

### **Panel Regression of Macroeconomic Indicators on U.S. Stock Returns**

Jun 2022 – Mar 2023

*Advisor: Prof. Saad Mouti*

- Reviewed literature on macroeconomic drivers of stock performance; applied robust and panel regression models in Python to evaluate predictive significance of inflation, interest rates, and GDP growth on cross-sectional equity returns.

## **Professional Experience**

---

### **SDIC Securities Co., Ltd.**

Shanghai, China

Sep 2023 – Apr 2024

#### *Derivatives Analyst Intern, Equity Derivatives Desk*

- Backtested options and futures trading strategies using Python (NumPy, pandas) and Excel VBA, analyzing historical P&L, Sharpe ratios, and maximum drawdowns to optimize entry/exit signals and improve risk-adjusted returns.
- Developed automated daily reporting system integrating real-time market data feeds with position tracking, reducing manual processing time by 60% and ensuring accurate Greeks calculation for portfolio risk management.
- Monitored trading positions across equity index options and stock index futures; proposed volatility arbitrage strategy improvements based on implied vs. realized volatility analysis.
- Collaborated with senior traders on structured product pricing and hedging strategies; assisted in client presentation materials for OTC derivatives solutions.

### **Taxpanda Inc.**

New York, USA

Jun 2022 – Aug 2022

#### *Data Analyst Intern*

- Cleaned and processed large-scale SAP-extracted financial datasets ; performed data quality checks and reconciliation to ensure accuracy for tax compliance reporting.
- Designed interactive Tableau dashboards for executive reporting, visualizing key metrics including revenue trends, expense breakdowns, and tax liability projections across multiple client portfolios.
- Collaborated with domain experts and clients on data collection methodology; documented ETL pipelines and created user guides for dashboard interpretation.

## **Test & Certifications**

---

- GRE General Test:** 332 (Verbal 162 + Quant 170)

Aug 2025

- CITI Program:** FDA-Regulated Research; Human Subjects Protection Biomed

Feb 2025

- IBM on Coursera:** Python for Data Science, AI & Development; Databases and SQL for Data Science with Python

2025 – Present

- Google on Coursera:** Data Analysis with R Programming

## **Activities & Honors**

---

- Member, Columbia University LEAP Center (Learning the Earth with Artificial Intelligence and Physics)

Aug 2025

- Champion, Penn Cup Soccer Tournament

Mar 2025

- Captain, CCST Future Soccer Team; Committee Member, Columbia Chinese Soccer Team

2025 – Present

- Vice Captain, UCSB Chinese Soccer Team

Jun 2021 – Jun 2023

## **Skills**

---

**Programming:** Python (NumPy, pandas, scikit-learn, PyTorch, SciPy), L<sup>A</sup>T<sub>E</sub>X, Git, SQL, Excel VBA, R.

**Machine Learning & Analysis:** Supervised Learning (Logistic Regression, XGBoost, Elastic Net, Cox Regression), Dimensionality Reduction & Clustering (PCA, Factor Analysis, UMAP, t-SNE, K-means, Leiden, Consensus Clustering), Time-Series & Causal Analysis (ARIMA, VAR, Granger Causality, DoWhy), Deep Learning (CNN, Grad-CAM, SHAP, Neural ODE, Symbolic Regression via PySR/AI-Feynman).

**Finance:** Options Pricing (Black-Scholes, Greeks), Portfolio Optimization, Risk Management, Backtesting Frameworks.

**Languages:** Mandarin (native), English (near-native; lived and studied in the U.S. for seven years).