Shuo Yin

Phone: (+86) 15690129730 | Email: yins25@mails.tsinghua.edu.cn | Homepage: https://yinshuo-thu.github.io/

EDUCATION

Tsinghua University, School of Economics and Management **Beijing, Sept 2025-Jun 2027**
Master of Finance
- **Current Courses**: Large Language Models and Generative AI, Financial Derivatives, Financial Data Analysis, Introduction to FinTech

Shanghai Jiao Tong University, School of Biomedical Engineering **Shanghai, Sept 2021-Jun 2025**
Bachelor of Engineering *(Electronic & Computer Eng. Track)*
- **GPA**: 3.8/4.0 | Merit Student Award | First Class Academic Scholarship | Outstanding Graduate
- **Core Courses**: Data Structure (A), Methods in Mathematical Physics (A), Biostatistics (A-), Medical Robot Control (A)

INTERNSHIP EXPERIENCES

Lingjun Investment (AUM ¥40B Chinese Quant) ***Quantitative Research Intern*** **Beijing, Aug 2025-Present**
- **Machine Learning Cross-Sectional Model:** Built LightGBM and MLP models to predict cross-sectional returns
  - **LightGBM:** LightGBM with 1k+ Alpha factor library for A-shares; integrate GAM for nonlinear main effects and EFB for high-dimensional interactions; 2021–2023 backtest: top 30% long-only annualized **42.0%**, Sharpe 2.25, MaxDD -8.8%
  - **MLP:** MLP with TabNet-style feature attention and adaptive DropPath for overfitting prevention; hyperparameter tuning (epochs, LR, architecture); 2021–2023 backtest top 30% long-only annualized **48.2%**, Sharpe 2.48, MaxDD -10.5%; live trading AR >**40%**
- **Deep Learning Time Series Strategy**: Developed GRU and Transformer baselines, with MoE combination attempts
  - **GRU:** Nonlinear input projection (linear mapping, LayerNorm, GELU); residual bidirectional GRU encoding, temporal attention aggregation, and MLP output; optimized loss. Backtest annualized return **57%**—top performer among strategies
  - **Transformer**: Linear projection with sinusoidal-cosine positional encoding; Encoder-Only architecture with ProbSparse self-attention and multi-scale decomposition; mixed-precision training and gradient norm clipping. Baseline backtest AR 44%.

Ant Group (Alipay) ***LLM Algorithm Intern*** **Shanghai, Apr 2025-Aug 2025**

Engineered ***ProQ*** model for automated data screening on *Ling LLM*; fine-tuned quality-filtering LLMs and streamlined the pipeline
- **Data Screening Model**: Built a custom agent-based data screening and harmonization model (***ProQ***), pre-screening data using metrics like PPL and iteratively selecting representative data for fine-tuning quality-filtering models. Post-trained *Ling MOE Lite* on ~3M optimized data, improving language understanding and code completion by over **1%**
- **Signal-Based Data Filters**: Developed rule- and model-based operators to identify low-quality data using targeted signals

Panoramic Hills Capital (AUM $4B Hedge Fund) ***Crypto Research Intern*** **Shanghai, Jan 2025-Mar 2025**

Conducted crypto market monitoring and developed sentiment indices on crypto, providing insights for fund positioning
- **NLP Sentiment Index Modeling:** Engineered quantitative sentiment indicator; curated datasets from X/Crypto News, LoRA-fine-tuned FinBERT on 8k labeled samples (cross-entropy loss, F1=0.91), yielding normalized daily scores [-1,1] to guide BTC trading
- **Crypto Analysis:** Monitored daily crypto trends; constructed financial models for BTC miners (MARA, CLSK, RIOT, etc.)

RESEARCH & PUBLICATION

AlphaMaster: Comprehensive Alpha Discovery and Agent Implementation via Hybrid Algorithms **Sep 2025-Present**

Applied machine learning–based quantitative modeling methods to refine equity strategies (e.g., MLP, GRU, Transformer)
- **Multi-source factor generation:** Integrated Agents, Genetic Algorithms, and Flow Networks to produce diversified alpha factors
  - **Flow Networks:** Added a structure-aware encoder with RGCN to capture mathematical relations; used flow-network exploration with dense, multi-aspect rewards to generate diverse, stable, and efficient alpha candidates
  - **Agent pipeline:** Generated interpretable, decay-resistant factors using multi-agent systems; implemented and validated efficacy
  - **Genetic algorithm:** Encoded factor expressions and applied crossover/mutation; selected robust factors using quantitative metrics
- **Factor alignment & Alpha reports:** Developed a factor alignment and filtering framework for A-shares and cryptocurrencies, implementing data cleansing, dual-chain(Generation and Evaluation Chain) dynamic iteration, and agent-based reporting

TS-Agent: RL Empowered LLM Agents for Financial Time Series Forecasting **Jun 2025-Sept 2025**
- Developed TS-Agent, a reinforcement learning–based agent for financial time-series forecasting, using an LLM-generated strategy pool for exploration-enhanced fine-tuning, then seeding stepwise RL from selected strategies with a custom RL module
- Improved strategy generation beyond baseline by over **10%**, with strategy performance close to top-tier LLMs (e.g., DeepSeek-R1)

Optiver-Trading at the Close (*Bronze Medalist*) **Sep 2023-Mar 2024**
- Derived high-frequency market microstructure factors from order book and auction data, including buy-sell imbalance ratios, price-volume triple barriers, reference price lags/differentials, statistical moments, depth-weighted spreads, and volatility proxies
- Built LightGBM model with rolling training and Optuna-based hyperparameter optimization; achieved MAE of 5.473 on validation set

Multimodal Large Models Paper(Under Review at Top Conferences)
- **ICLR 2026, First Author |** BagelScore: Visual-Language Evaluation Made Easy
- **CVPR 2026, Co-First Author |** OGDiff: Rethinking Open Set Domain Generalization: A Conditional Diffusion Perspective

ADDITIONAL INFORMATION
- **Language:** English (TOEFL: 108), Chinese (Native)
- **Skills:** Python, C++, SQL; Git, Latex; Wind, Excel/PowerPoint
- **Hobbies:** Basketball(Vice Captain of College Team, Center/PF), regular fitness training, Go and Tennis

尹硕

电话：15690129730 | 邮箱：yins25@mails.tsinghua.edu.cn | 个人主页：https://yinshuo-thu.github.io/

教育经历

清华大学 | 经济与管理学院 | 金融硕士 2025/9 – 2027/6

- **在修课程**：大模型与生成式人工智能，金融衍生工具，金融数据分析，金融科技导论

上海交通大学 | 生物医学工程学院 | 工学学士(电子信息方向) 2021/9 – 2025/6

- GPA: 3.8/4.0 | 上海交通大学 A 等学业奖学金 | 上海交通大学优秀毕业生
- **专业课程**：数据结构，计算机辅助手术，数理方法，信号与系统，生物统计

实习经历

灵均投资，量化策略实习生（深度学习模型） 2025/8 – 至今

- **机器学习截面多因子选股**：基于 LightGBM 和 MLP 等机器学习模型实现多因子截面收益率预测，进行系统调优与回测
  - **树模型**:LightGBM 结合 A 股 1k+ Alpha 因子库构建模型，集成 GAM 捕捉非线性主效应并采用 EFB 压缩高维交互，滚动训练关键交易日预测，21-23 年回测前 30%多头年化 **42.0%**、夏普 2.25、MaxDD-8.8%；组合优化后 2025 年回测 **30%+**
  - **MLP**：实现多层感知机(MLP)深度学习量化策略，嵌入 TabNet 式特征注意力机制与自适应 DropPath 防过拟合，系统调参（轮次、LR、网络结构）；21-23 年回测前 30%多头年化 **48.2%**、夏普 2.48、MaxDD-10.5%；近期实盘年化 **40%+**
- **深度学习时序策略**：建立时间序列数据输入，完成 GRU 和 Transformer 完整基线策略，进行 **MoE** 组合尝试
  - **GRU**:通过输入投影层对原始特征进行非线性变换（线性映射、LayerNorm 归一化和 GELU 激活），残差双向 GRU 编码后进行时序注意力聚合和 MLP 输出，优化损失函数，回测年化收益提高至 **57%**，多策略中表现最佳
  - **Transformer**:在线性层映射后进行正弦-余弦位置编码，采用 Encoder-Only 架构结合 ProbSparse 高效自注意力和多尺度时序分解，使用混合精度计算加速收敛并对梯度进行范数裁剪，基线回测年化收益 **44%**

蚂蚁集团（支付宝），大模型算法实习生 2025/4 – 2025/8

- **数据和模型优化**：基于 Agent 和大模型微调，从头建立问题筛选和指标归纳模型 ***ProQ***，以筛选、预测低质量和问题数据；通过 PPL 指标预筛选、迭代归纳问题、微调特异打标小模型的流程批量识别问题数据，服务百灵 MOE 等大模型
- **框架验证**：通过 ***ProQ*** 框架在约 3M 代码、合成数据上后训练测试，语言理解、代码补全等能力上实现 **1%**以上提升
- **高阶算子**：建立高阶通用算子，归纳因子信号、结合模型筛选低质量训练数据，参与百灵 Ling 和 Ring 系列模型基座优化

Panoramic Hills Capital，加密货币实习生 2025/1 – 2025/3

- **NLP 情绪指数建立**：独立设计量化情绪指标，从 X/Crypto 新闻网站等源采集文本，LoRA 微调 FinBERT 模型（8k 标注样本，交叉熵损失，F1=0.91），输出标准化日频情绪分数[-1,1]，作为技术指标指导 BTC 基本面交易仓位动态调整
- **Crypto 研究**：跟踪加密市场动向并总结信息流，给出归纳和仓位指导；建立 MARA、CLSK 等主要 BTC 矿企业务财务模型

项目&竞赛

AlphaMaster：多算法源下的因子挖掘与 Agent 评估筛选框架 2025/9 – 至今

- **多源算法因子生成**：整合 Agent、遗传算法、生成流网络等算法构建多元化因子，覆盖数据驱动与逻辑驱动的挖掘能力
  - 生成流网络：结构感知编码器基于 RGCN 捕捉数学结构，通过密集多奖励结构提供丰富反馈，提供多样化稳定因子生成
  - Agent：通过多 Agent 实现可解释且抗衰减 Alpha 因子挖掘，结合 Qlib 项目完成假设生成、构建、回测得到有效因子
  - 遗传算法：将因子表达式编码后进行交叉变异生成，以 IC 值、收益率等指标为适应度函数筛选最优稳健 Alpha 因子
- **因子筛选和报告框架**：在 A 股和加密货币上实现对齐、双链迭代和 Agent 报告，AR、IR 等指标和效率优于其他 LLM 方法

Optiver-Trading at the Close | 铜牌 2023/9 – 2024/3

- 基于订单簿和拍卖期数据，挖掘高频市场微观结构因子，包括买卖不平衡度、价量三重比率、参考价位滞后/差分、统计矩、订单深度加权价差及波动率代理，结合时序差分与滚动窗口聚合，实现因子有效性验证与噪声过滤
- 构建 LightGBM 预测模型，时序滚动训练并使用 Optuna 进行超参数优化，重训练后在验证集上 MAE 为 **5.473**

期刊&顶会

TS-Agent：基于强化学习的金融时间序列预测 Agent 框架 | B 会一作&通讯 2025/6 – 2025/9

- 建立时间序列预测 Agent 框架，使用 GLM/GPT5 生成策略轨迹，对 Qwen7B 进行增强微调,生成策略效果提升 **10%+**
- 建立策略池后随机起始，通过专用奖励进行分步 PPO 强化学习，验证集预测策略生成能力追平 Deepseek R1 等基线大模型

LNG 现货运费的多重分形特征：洞见、预测与交易策略 | SCI 二区共一 2024/2 – 2024/9

- 基于 MF-DFA 等计算和多峰分析验证 LNG 现货运费多重分形性，建立因子信号(谱分解成分和峰值位置、长程相关性强度等)
- 使用 LSTM 等方法进行 LNG 期货收益预测，结合 Hurst 指数等指标实现策略切换并进行窗口平滑，回测年化收益率超 **36%**

多模态在投一作顶会

- **ICLR2026 一作| BagelScore**：基于统一多模态模型的视觉-语言评估范式
- **CVPR2026 共一| OGDiff**：条件扩散驱动的开放集域泛化优化框架

其他信息

语言能力：托福 108，普通话二甲

工作技能：Python（Pytorch），SQL，C++；Git，Latex；Office,Wind/Bloomberg

兴趣爱好：篮球（院队中锋/大前锋）,围棋，网球