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**: 3.75/4.0 | 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)
- **Research**: Multimodal LLM(AAAI, ICLR); Financial Time Series (SCI Q2 journal); Agent, Reinforcement Learning (CCF-B conf.)

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:** Trained on 1k+ alpha factors with rolling windows to forecast cross-sectional returns; 2021–2023 backtest (long top-30%) hit **42%** AR, Sharpe 2.25, max DD −8.8%; 2025 live-simulated trading, AR **30%+** after portfolio optimization
  - **MLP:** Designed and tuned an MLP (epochs, learning rate, network) with rolling predictions; 2021–2023 backtest top-30% long portfolio achieved AR **48.2%**, Sharpe 2.48, max drawdown −10.5%; recent live-simulated trading shows **40%+** annualized
- **Deep Learning Time Series Strategy**: Developed GRU and Transformer baselines, with MoE combination attempts
  - **GRU:** Added an input-projection block (Linear → LayerNorm → GELU), followed by a bidirectional GRU encoder and temporal-attention pooling, with an MLP prediction head; loss-driven optimization improved backtest AR by 2%+ to **57%**
  - **Transformer**: Implemented an encoder-only Transformer with linear feature mapping and sinusoidal positional encoding; currently developing multi-scale temporal fusion and attention refinements; 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:** Built scrapers and real-time pipelines to source, aggregate, and filter crypto feeds; fine-tuned a BERT sentiment model on the collected data to produce an index used to guide position sizing
- **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　　**Aug 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:** Employed multi-agent discovery for interpretable, decay-resistant factors; leveraged Qlib for hypothesis formation, factor construction, and walk-forward backtesting to validate 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)

Kaggle Competition: Optiver-Trading at the Close (*Bronze Medalist*)　　　　　　　　**Sep 2023-Mar 2024**
- Built a LightGBM-based predictive model, forecasting Nasdaq stock closing price movements from order book and auction data
- Implemented a purged time-series cross-validation to mitigate lookahead bias, applied hyperparameter tuning with early stopping, and retrained a final ensemble model, achieving a Mean Absolute Error (MAE) of **5.473** on validation

ADDITIONAL INFORMATION

- **Language：** English (TOEFL: 108), Professional Working Proficiency
- **Technical Skills:** Python, C++; SQL; Git, Wind, Excel/PowerPoint
- **Hobbies:** Basketball(Vice Captain of College Team, Center/PF), Go and Tennis

<h1 style="text-align:center">尹硕</h1>

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**：3.75/4.0｜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
- **研究经历**：多模态大模型(AAAI, ICLR在投)，金融时间序列(SCI二区发表)，Agent和强化学习（CCF-B发表）

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：基于LightGBM结合A股1k+ Alpha因子构建模型，预处理和滚动训练对关键交易时点进行截面收益率预测，21-23年回测前30%多头年化回报率 **42.0%**，夏普比率2.25，最大回撤-8.8%，组合优化后2025年回测收益率**30%+**
  - MLP：实现多层感知机(MLP)深度学习量化策略，系统调参（轮次、学习率、网络结构）并进行滚动预测，21-23年回测前30% 多头年化回报率 **48.2%**，夏普比率 2.48，最大回撤 – 10.5%；近期实盘回测年化收益**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），进行残差连接的双向GRU编码后进行时序注意力聚合，通过多层感知机输出预测，并通过损失函数优化将回测年化收益提高**2%+**至57%
  - Transformer：在线性层映射后进行正弦-余弦位置编码，采用Encoder-Only架构，正在进行多尺度时序融合和注意力机制优化，启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，通过因子或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，加密货币研究实习生　　　　　　　　　　　　　　　　　　　　　2025/1－2025/3
- **NLP情绪指数建立**：独立完成情感指数的数据收集、部署、训练全流程，归纳社交平台和媒体信息源，实现代理库实现爬虫，标记后微调Bert模型建立市场情绪判断模型，与FireChart等其他信号被一并用于仓位指导和调整
- **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矿企财务模型

研究&项目

AlphaMaster：多算法源下的因子挖掘与Agent评估筛选框架　　　　　　　　　　　　　　　　　2025/8－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关系图卷积网络（RGCN）捕捉数学结构，借助流网络特性实现更稳定的探索，通过密集多方面奖励结构提供丰富反馈，实现多样化、高效且稳定的Alpha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
- **因子对齐和筛选报告框架**：在A股和加密货币上实现清洗对齐、双链动态迭代和Agent报告，生成链后，优化链进行回测反馈和相似度评估，CSI500/1000基准测试中年化收益率、IR等指标和效率优于传统方法与其他LLM-based方法

TS-Agent：基于强化学习的金融时间序列预测Agent框架 | 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%+**
- 建立策略池后取随机起始点，通过专用奖励模块进行分步PPO强化学习，在已知/未知测试任务上的预测策略生成能力追平Gemini 2.5Pro、GLM4.5、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 | 共同一作　　　　　　　　　　　　　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%**

Kaggle竞赛：Optiver-Trading at the Close | 铜牌(前10%)　　　　　　　　　　　　　　2023/9－2024/3
- 构建了LightGBM预测模型，利用订单簿和拍卖数据预测纳斯达克股票收盘价的波动，并通过特征工程提取了市场微观结构特征（如不平衡度、三重比率、滞后/差异、统计矩等），实现滚动时序交叉验证以进行调优与验证
- 进行超参数调优并使用早停策略，重训练得到最终模型，在验证集上平均绝对误差（MAE）**5.473**

语言&技能

英语能力：托福108（听力30 阅读30 口语23 写作25）；普通话二甲

计算机技能：Python（Pytorch），SQL，C++；MS Office；Wind/Bloomberg

兴趣爱好：篮球（院队中锋/大前锋），围棋，网球