Yuncong Liu

(+86)15600198337 yuncongliu0703@gmail.com www.linkedin.com/in/ycliu0703 https://github.com/ycsama0703

Education

National University of Singapore **Aug 2025 – Jun 2029(Expected)**
PhD in Digital Financial Technology *Singapore*
Research Focus: Multimodal LLMs for finance, Trustworthy machine learning and deep learning

Columbia University **Sep 2023 – Feb 2025**
Master of Arts in Statistics, GPA: 3.9/4.0 *New York, USA*
Machine Learning, Applied Data Science, Advanced Probability Theory, Statistical Machine Learning

Central University of Finance and Economics **Sep 2019 – Jun 2023**
Bachelor of Science in Actuarial Science, GPA: 88/100 *Beijing, China*
Mathematical Analysis, Advanced Algebra, Financial Mathematics, Multivariate Statistical Analysis, Stochastic Processes

Publications

Yixuan Liang, **Yuncong Liu**, Boyu Zhang, Christina Dan Wang, Hongyang Yang.
"FinGPT: Enhancing Sentiment-Based Stock Movement Prediction with Dissemination-Aware and Context-Enriched LLMs."
1st Workshop on Preparing Good Data for Generative AI: Challenges and Approaches @ AAAI 2025.
arXiv:2412.10823 [cs.CL]
https://doi.org/10.48550/arXiv.2412.10823

Awards and Recognitions

National Olympiad in Informatics (NOIP 2016): Second Prize (Top 50 in Province)
Central University of Finance and Economics Outstanding Scholarship 2020: Second Class (top 6%)
Mathematical Contest in Modeling 2021: Honorable Mention
Munich Re Actuarial Mathematics Competition 2022: Award of Excellent Standing (Global Top 50)
Munich Re Actuarial Mathematics Competition 2022: First Place in Problem-Solving Section

Research Experience

FinGPT: Open-Source Financial LLM | *Research Assistant* **Jun 2024 – Jan 2025**

- Constructed financial domain-specific corpora using FinHub API, integrating structured data like SEC filings and earnings call transcripts. Built automated scraping and cleaning pipelines with Scrapy and BeautifulSoup.
- Led secondary pretraining and instruction fine-tuning of financial LLMs, incorporating temporal attention mechanisms and domain-adaptive training strategies. Responsible for fine-tuning specialized knowledge base models.
- Configured LoRA adapters for attention modules and FFN layers during model lightweighting, achieving parameter-efficient updates via rank decomposition matrices. Implemented dynamic quantization strategies to optimize inference efficiency.

Huaxi Securities Research Institute | *Quantitative Research Assistant* **Nov 2021 – Mar 2022**

- Developed Python implementation of Barra multi-factor system based on CNE5 risk model, designing efficient numerical computation frameworks to optimize factor exposure calculation.
- Built automated data pipelines from Wind API to local models, creating PySpark-based distributed factor data preprocessing modules supporting real-time computation and validation.
- Participated in LSTM-based RNN stock selection strategy research, developing experimental platforms in Python and achieving 24.33% annualized return in backtesting.

Machine Learning-Based Campus Vending Machine Valuation | *Research Team Lead* **Feb 2021 – May 2021**

- Developed dual-model analytical framework combining GLM and random forest, conducting feature permutation importance analysis to identify core profitability factors.
- Built A/B testing platform using Python Flask, designing QR-code electronic coupon systems to quantify advertising impact on consumer behavior.
- Constructed LSTM time-series prediction models to capture semester-cycle sales fluctuations, providing raw material supply recommendations to distributors.

Professional Experience

AI4Finance **Jun 2024 – Aug 2024**
Financial NLP Research Assistant *New York, USA*

- Built and maintained data pipelines from financial information sources including research reports and earnings calls, ensuring sufficient training data while reducing noise impact through feature selection.
- Utilized NLP tools (NLTK, SpaCy) to clean financial text data and generated context-rich vector representations using BERT models to capture financial domain-specific language patterns.
- Designed sentiment analysis pipeline employing BERT and LSTM models for fine-grained classification, incorporating sentiment features into multimodal datasets to enhance market trend prediction.

AIG (American International Group) **Jun 2024 – Aug 2024**
Data Science Intern *New York, USA*

- Developed CNN-based customer physiological age prediction system using ResNet-34 with spatial attention mechanism, designing DICOM medical image preprocessing pipelines to standardize resolutions.
- Built PCA-GLM framework with VIF-based feature selection, optimizing lapse rate prediction through principal component analysis and LASSO regression cross-validation.
- Designed optimal binning algorithms using KS tests to transform continuous age variables into risk intervals, implementing sample weighting mechanisms to address data imbalance.

Munich Re **Aug 2022 – Jan 2023**
Data Engineering Intern *Beijing, China*

- Designed high-performance SQL ETL processes for heterogeneous data from 12 insurers, developing data consistency validation modules using window functions and recursive CTEs.
- Redesigned SQL accounting models through query optimization, improving performance for complex actuarial calculations including IBNR reserves. Created VBA data loaders for Excel-SQL Server interaction.
- Built R-SQL hybrid analysis framework utilizing random forest models to identify high-claim risk patterns.

Alibaba Group **Mar 2022 – Jul 2022**
Business Analysis Intern *Beijing, China*

- Developed Python-based order conversion rate monitoring models through user behavior and market data analysis, optimizing pandemic-impacted homestay business workflows.
- Conducted EDA and analytical tasks using Python and R, creating over 30 data visualizations and tables for logistic regression, decision tree, and random forest models.

Technical Skills

Programming Languages: Python, C++, C, R, EXCEL VBA, SQL
Developer Tools: VS Code, Pycharm, Google Cloud Platform, Jupyter Notebook
Libraries/Frameworks: Pandas, Numpy, Pytorch, Matplotlib, Seaborn, Scikit-learn, XGBoost