Zhang Yifei

Gulou Campus, Nanjing University, No.22 Hankou Road, Gulou District, Nanjing, Jiangsu Province, China |
(+86)182-5123-9810 | hoderzyf@gmail.com | Postcode: 210093 | HomePage: hoder-zyf.github.io

Education

Department of Computer Science and Technology — Nanjing University
B.Sc. in Computer Science and Technology (FinTech) ***Top Talent Class*** — *Sep 2020 – Jun 2024*

School of Management & Engineering — Nanjing University
M.Sc. in Financial Engineering and FinTech — *Sep 2024 – Jun 2027*
Supervisor: *Prof.* ***Honghai Yu***

INDUSTRY INTERN EXPERIENCE

- **Nanjing Securities** — *June 2022 – July 2022*
  Software Development Engineer at Fintech Dept.
- **Deloitte** — *Oct 2022 – Dec 2022*
  Financial Report Smart Evaluation Intern at FengYu Intelligent Technology
- **Soochow Securities** — *Oct 2023 – Jan 2024*
  Intern in Retail, Social Services, and Beauty Care Group

ACADEMIC INTERN EXPERIENCE

- **The Chinese University of Hong Kong, Shenzhen (CUHK-SZ)** — *July 2024 – Present*
  Research intern advised by Prof. ***Benyou Wang***

Research & Projects

IFLYTEK Spark Camp — *Aug 2023 – Aug 2023*
Legal Pleading Generation System based on Spark 2.0 Model — *Project Member*

- Conducted a comparative analysis of **GPT-4** and **Spark 2.0**, examining their **architectures, functionalities, and potential applications** across diverse domains, ultimately showcasing **Spark 2.0's superior contextual understanding and content generation capabilities**.
- Developed a **Spark 2.0-based legal pleading generation system** designed to automate legal document creation, employing **advanced NLP techniques** to analyze case details and generate coherent, legally compliant pleadings.
- Implemented **custom training and fine-tuning strategies** for the Spark 2.0 model using a specialized legal document dataset, significantly improving the model's accuracy and relevance in generating contextually appropriate legal text.
- Evaluated the system's performance through comprehensive testing with real-world legal scenarios, achieving **significant reductions in document preparation time** alongside enhancements in **legal pleading quality and precision**.

National Undergraduate Innovation Training Program — *Dec 2022 – Dec 2023*
Volatility Prediction by Introducing Investor Sentiment — *Project Member*

- Developed a multi-faceted approach to constructing predictive factors for market volatility by integrating diverse indicators, including mutual fund clustering, **Clique-based analysis for identifying institutional herding**, **refined LSV indicators**, implied volatility metrics, and the Baidu search index. This strategy aimed to holistically capture investor sentiment and enhance prediction accuracy.
- Engineered a **hybrid LSTM-autoencoder model** for volatility prediction, selected after a rigorous evaluation of various machine learning models. This architecture effectively processes time-series data and captures the dynamic nature of investor sentiment. Addressed critical challenges during model development, including hyperparameter optimization, imbalanced data handling, and robust training and testing protocols, culminating in successful real-world data prediction.

Undergraduate Thesis (Outstanding) *Dec 2023 – May 2024*

Investor Behavior Consistency on the Xueqiu Platform: A Textual Analysis *Author*

- Utilized empirical data from Xueqiu, a leading social trading platform, to investigate the alignment between investor sentiment, as expressed in online discourse, and actual trading behavior. Employed **NLP techniques**, **sentiment dictionaries**, **BGE-M3 embeddings**, and **GPT-4** for text analysis and sentiment extraction.
- Performed a comprehensive analysis, revealing a significant correlation between investor narratives on Xueqiu and their corresponding market actions. Identified key factors influencing this consistency, including investor gender and the information context, using variables such as investor profile information and clarity of expressed sentiment.
- Assessed the market impact of this investor behavior consistency by calculating **Cumulative Abnormal Returns (CAR)** and **Investment Returns (Inv_ret)**, demonstrating the tangible influence of social media sentiment on financial market dynamics.

Financial MultiModal LLMs June 2024 – Oct 2024

Advised by Prof. Benyou Wang, Prof. Honghai Yu and Jimin Huang *Project Leader*

- Developed and released **FinLLaVA-8B**, a multimodal Large Language Model (LLM) for financial applications, achieving an MMMU (Overall) score of 36.3 and an MMMU (Business) score of 30.7. This involved creating a **specialized financial multimodal benchmark dataset** for training and evaluation, and fine-tuning the LLM on text, charts, tables, and numerical data, aligning it with real-world financial scenarios.
- Designed and constructed the **UCFE (User-Centric Financial Expertise)** benchmark, a novel framework for evaluating LLM performance in real-world financial applications, featuring a purely text-based, multi-turn dialogue setting and a user simulator to mimic realistic user interactions.

Large-scale AI Stock Market Simulation *Oct 2024 – Present*

Advised by Prof. Benyou Wang and Prof. Honghai Yu *Project Leader*

- Developing a large-scale, Agent-Driven simulation framework ——**TwinMarket** using LLM agents to model investor behavior, aiming to validate economic theories within China's financial context.
- Building scalable simulations, inspired by the **Santa Fe Institute** and **Stanford's Artificial Town**, to analyze collective trading's impact on market outcomes and provide insights for **policy making and China's economic development**.

Publications

Open-FinLLMs: Open Multimodal Large Language Models for Financial Applications
Qianqian Xie, Dong Li, Mengxi Xiao, Zihao Jiang, Ruoyu Xiang, Xiao Zhang, Zhengyu Chen, Yueru He, Weiguang Han, Yuzhe Yang, Shunian Chen, **Yifei Zhang**, et al. *arXiv preprint 2408.11878, 2024.*

UCFE: A User-Centric Financial Expertise Benchmark for Large Language Models
Yuzhe Yang*, **Yifei Zhang***, Yan Hu, Yilin Guo, Ruoli Gan, Yueru He, Mingcong Lei, Xiao Zhang, Haining Wang, Qianqian Xie, Jimin Huang, Honghai Yu, and Benyou Wang *arXiv preprint 2410.14059, 2024.*, **NAACL Findings 2025** ; **# 1 Paper of the day on Hugging face**

Do investors' actions speak louder than words?
Honghai Yu, Zhuo Chen, Yunmiao Zhang, Haining Wang, and **Yifei Zhang**, ***An extended version of my undergraduate thesis,*** **Accepted by the 21st Annual Conference on Financial Engineering and Risk Management**

TwinMarket: A Scalable Behavioral and Social Simulation for Financial Markets
Yuzhe Yang*, **Yifei Zhang***, Minghao Wu*, Kaidi Zhang, Yunmiao Zhang, Honghai Yu, Yan Hu, and Benyou Wang *arXiv preprint 2502.01506, 2025.*

Honors and Awards (Selected)

Province First Prize, National College Students English Vocabulary Ability Competition (**top 5%**)	*May 2021*
People's Scholarship of Nanjing University	*Oct 2021 & Nov 2022*
Elite Scholarship, **Top Talent Class** of Nanjing University (**top 30%**)	*Nov 2021*
Third Prize, 2022 National Student Data Analysis Competition (**top 30%**)	*Dec 2022*
Excellence Scholarship, **Top Talent Class** of Nanjing University (**top 15%**)	*Dec 2023*
Second Prize, Third "Xueshi Cup" Academic Paper Competition of Nanjing University (**top 15%**)	*May 2024*