Yuzhe Yang

TD413, 2001 Longxiang Avenue,
Longgang District, Shenzhen, 518172, China

(+86) 183-1076-2536
yuzheyang@link.cuhk.edu.cn
tobyyang7.github.io
google scholar

Education

The Chinese University of Hong Kong, Shenzhen — Shenzhen, China
B.Eng. in Computer Science and Engineering — *Sep. 2021 – May 2025*

Selected Courses

Natural Language Processing *(PhD Level)*, Large Language Model *(PhD Level)*, Parallel Programming, Design and Analysis of Algorithms, Operation Systems, Computer Architecture, Optimization, Data Structures

Publications

[1] **UCFE: A User-Centric Financial Expertise Benchmark for Large Language Models**

Yang, Y., Zhang, Y., Hu, Y., Guo, Y., Gan, R., He, Y., Lei, M., Zhang, X., Wang, H., Xie, Q., Huang, J., Yu, H., Wang, B., *arXiv preprint 2410.14059*. 2024, (NAACL Findings 2025). [link]

[2] **FAST-CA: Fusion-based Adaptive Spatial-Temporal Learning with Coupled Attention for airport network delay propagation prediction**

Li, C., Qi, X., Yang, Y., Zeng, Z., Zhang, L., Mao, J., *Information Fusion*. 2024. P. 102326

Preprints

[3] **TwinMarket: A Scalable Behavioral and Social Simulation for Financial Markets**

Yang, Y., Zhang, Y., Wu, M., Zhang, K., Zhang, Y., Yu, H., Hu, Y., Wang, B., *arXiv preprint 2502.01506*. 2025, (ICLR 2025 Workshop on Advances in Financial AI). [link]

[4] **Open-FinLLMs: Open Multimodal Large Language Models for Financial Applications**

Xie, Q., Li, D., Xiao, M., Jiang, Z., Xiang, R., Zhang, X., Chen, Z., He, Y., Han, W., Yang, Y., Chen, S., Zhang, Y., Shen, L., Kim, D., Liu, Z., Luo, Z., Yu, Y., Cao, Y., Deng, Z., Yao, Z., Li, H., Feng, D., Dai, Y., Somasundaram, V., Lu, P., Zhao, Y., Long, Y., Xiong, G., Smith, K., Yu, H., Lai, Y., Peng, M., Nie, J., Suchow, J. W., Liu, X.-Y., Wang, B., Lopez-Lira, A., Huang, J., Ananiadou, S., *arXiv preprint 2408.11878*. 2024. [link]

[5] **FedDTPT: Federated Discrete and Transferable Prompt Tuning for Black-Box Large Language Models**

Wu, J., Chen, S., Yang, Y., Li, Y., Hou, S., Jing, R., Wang, Z., Chen, W., Tian, Z., *arXiv preprint 2411.00985*. 2024. [link]

Research Experience

Large-scale Social Simulation Agent — Oct. 2024 – Jan. 2025
Research Assistant Internship, advised by Prof. Benyou Wang and Honghai Yu — *CUHK-Shenzhen*

- Developed the TwinMarket framework where LLM agents simulate large-scale human investor behaviors to validate economic principles and market theories, leveraging the TwinMarket framework to model individual decision-making and social interactions.
- Built scalable simulations to evaluate how collective trading behavior impacts broader market outcomes, successfully replicating key market phenomena such as volatility clustering and fat-tailed return distributions.
- This work had submitted to *ICML 2025* [3]

Financial Multimodal LLM — May. 2024 – Oct. 2024
Research Assistant Internship, advised by Prof. Benyou Wang and Jimin Huang — *CUHK-Shenzhen*

- Leaded a task team to the multimodal extension of LLM; this work had submitted to *ARR Dec.* [4]
- Developed a multimodal financial benchmark dataset for LLM training and evaluation
- Multimodal instruction finetuning for LLM, include text, image (chart & tabular) and numerics data

- Align multimodal LLM with financial data and real-world scenarios to improve model performance
- Released FinLLaVA-8B: Achieved MMMU (Overall) score of 36.3 and MMMU (Business) score of 30.7
- Constructed a purely text-based multi-turn dialogue benchmark to evaluate the performance of LLMs in real-world financial applications using a user simulator; this work had accepted to *NAACL Findings 2025* [1]; #1 Paper of the day on Hugging face

Flight Delay Propagation Modeling Aug. 2023 – Sep. 2024

Undergraduate Research Assistant, advised by Prof. Jianfeng Mao *CUHK-Shenzhen*

- Developed a GNN framework integrating dynamic and adaptive graph learning with coupled attention mechanisms to address complex spatial-temporal dependencies in airport delay propagation; this work had published in *Information Fusion* [2]
- Enhanced the SIS epidemiological model by incorporating adaptive graph learning to simulate and predict epidemic transmission dynamics in airport networks; this work is in preparation [6], to be submitted to *Transportation Research Part B: Methodological*
- Leverages Neural ODE networks to improve flight delay prediction by developing a continuous graph model that enhances interpretability, reduces training time, and addresses challenges like irregular time sampling and missing data

Projects

Quant-GPT: Money is All You Need [link] | *PyTorch, Transformers, ChromaDB* Mar. 2024 – Apr. 2024

- Final project for the PhD course CSC6052, a multi-agent system for A-share market investment decisions
- Fine-tuned an LLM, integrating it with sentiment analysis and real-world market data.
- Utilized RAG and multi-agent systems to dynamically access and synthesize relevant financial news, enhancing the model's ability to forecast market trends and returns
- Results achieved: Sharpe Ratio: 0.40, Annualized Return: 7.26%, Max Drawdown: 13.61%

Travel Insurance Recommendation AI System [link] | *PyTorch, LangChain* Jan. 2024 – Apr. 2024

- Developed an AI system to predict flight delays and recommend personalized travel insurance, enhancing customer satisfaction
- Fine-tuned the LLM using an insurance corpus to improve domain-specific question-answering capabilities, achieving an 83% accuracy in identifying user intent
- Utilized deep learning and LLM agents for accurate delay predictions and customer sentiment assessment

Flight Information System [link] | *Python, LangChain, SQL, Flask* Mar. 2024 – Apr. 2024

- Developed database system to optimize airline management, including passenger bookings and flight logistics
- Delivered a functional database with a user-friendly web interface
- Integrated LLM to enhance database architecture and query generation

Work Experience

China Telecom Beijing Research Institute Jan. 2024 – Mar. 2024

Remote Internship *Beijing, China*

- Intern at the AI Large Model Research Team
- Analyze a technology's trends, applications, and industry impact

Shenzhen Branch of China Telecom Jan. 2024 – Apr. 2024

Part-time Internship *Shenzhen, China*

- Time Series Analysis, Data Visualization
- GIS Data Analysis, Data Mining

Technical Skills

Languages: Python, C/C++
Developer Tools: Git, Docker, VS Code, Linux
Libraries: PyTorch, Transformers

Service

Reviewer: IJCAI 2025, ICLR 2025 Workshop

Latest Version
Last updated on March, 2025