

Yuzhe Yang

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

☎ (+86) 183-1076-2536
✉ yuzheyang@link.cuhk.edu.cn
🌐 tobyayang7.github.io
🔗 [google scholar](#)

EDUCATION

The Chinese University of Hong Kong, Shenzhen
B.Eng. in Computer Science and Engineering

Shenzhen, China
Sep. 2021 – May 2025

SELECTED COURSES

Natural Language Processing (*PhD Level*), Large Language Model (*PhD Level*), Parallel Programming (*PhD Level*), 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] 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). [\[link\]](#)

[3] 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

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

Huang, J., Xiao, M., Li, D., Jiang, Z., Yang, Y., Zhang, Y., Qian, L., Wang, Y., Peng, X., Ren, Y., Xiang, R., Chen, Z., Zhang, X., He, Y., Han, W., Chen, S., Shen, L., Kim, D., Yu, Y., Cao, Y., Deng, Z., Li, H., Feng, D., Dai, Y., Somasundaram, V., Lu, P., Xiong, G., Liu, Z., Luo, Z., Yao, Z., Weng, R.-L., Qiu, M., Smith, K. E., Yu, H., Lai, Y., Peng, M., Nie, J.-Y., Suchow, J. W., Liu, X.-Y., Wang, B., Lopez-Lira, A., Xie, Q., Ananiadou, S., Tsujii, J., *arXiv preprint 2408.11878*. 2025. [\[link\]](#)

[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* [\[2\]](#)

Financial Multimodal LLM

May. 2024 – Oct. 2024

Research Assistant Internship, advised by Prof. [Benyou Wang](#) and [Jimin Huang](#)

CUHK-Shenzhen

- Led a task team to the multimodal extension of LLM; this work had submitted to *ARR Dec.* [\[xie2024openfinllmsopenmultimodallarge\]](#)
- 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
<i>Undergraduate Research Assistant, advised by Prof. Jianfeng Mao</i>	<i>CUHK-Shenzhen</i>
<ul style="list-style-type: none"> • 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 <i>Information Fusion</i> [3] • 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 <i>Transportation Research Part B: Methodological</i> • 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] <i>PyTorch, Transformers, ChromaDB</i>	Mar. 2024 – Apr. 2024
<ul style="list-style-type: none"> • 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] <i>PyTorch, LangChain</i>	Jan. 2024 – Apr. 2024
<ul style="list-style-type: none"> • 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] <i>Python, LangChain, SQL, Flask</i>	Mar. 2024 – Apr. 2024
<ul style="list-style-type: none"> • 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
<i>Remote Internship</i>	<i>Beijing, China</i>
<ul style="list-style-type: none"> • 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
<i>Part-time Internship</i>	<i>Shenzhen, China</i>
<ul style="list-style-type: none"> • 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