

# Yizhang ZHU

📞 +86 19936076638 📩 yzhu305@connect.hkust-gz.edu.cn 🏠 Homepage 🐧 Github

## EDUCATION

<b>The Hong Kong University of Science and Technology (Guangzhou)</b>	<b>2025.09 - Present</b>
Ph.D. Student in Data Science and Analytics	
Supervisor: Prof. Yuyu LUO	
<b>The Hong Kong University of Science and Technology (Guangzhou)</b>	<b>2023.09 - 2025.07</b>
M.Phil. in Data Science and Analytics	
Supervisor: Prof. Yuyu LUO	
<b>Chongqing University</b>	<b>2019.09 - 2023.06</b>
B.Eng. in Computer Science and Technology	

## EXPERIENCE

<b>Tsinghua University</b>	<b>2025.06 - 2025.09</b>
Visiting Student in Database Group, Department of Computer Science and Technology	
Supervisor: Prof. Guoliang LI	
<b>National University of Singapore Chongqing Research Institute</b>	<b>2022.09 - 2023.05</b>
Visiting Student in Computer Engineering Joint Program	
Supervisor: Prof. Yung Chii LIANG	

## PUBLICATIONS

<b>A Survey of Data Agents: Emerging Paradigm or Overstated Hype?</b>	<a href="#">Link ↗</a>
Yizhang ZHU, Liangwei WANG, Chenyu YANG, et al., Guoliang LI, Yuyu LUO	
Preprint (Survey)	
<b>EllieSQL: Cost-Efficient Text-to-SQL with Complexity-Aware Routing</b>	<a href="#">Link ↗</a>
Yizhang ZHU, Runzhi JIANG, Boyan LI, Nan TANG, Yuyu LUO	
Conference on Language Modeling (COLM 2025)	
<b>Are Large Language Models Good Statisticians?</b>	<a href="#">Link ↗</a>
Yizhang ZHU, Shiyin DU, Boyan LI, Yuyu LUO, Nan TANG	
Advances in Neural Information Processing Systems (NeurIPS 2024)	
<b>LEAD: Iterative Data Selection for Efficient LLM Instruction Tuning</b>	<a href="#">Link ↗</a>
Xiaotian LIN, Yanlin QI, Yizhang ZHU, Themis Palpanas, Chengliang CHAI, Nan TANG, Yuyu LUO	
International Conference on Very Large Data Bases (VLDB 2026)	
<b>RAMer: Reconstruction-based Adversarial Model for Multi-party Multi-modal Multi-label Emotion Recognition</b>	<a href="#">Link ↗</a>
Xudong YANG, Yizhang ZHU, Nan TANG, Yuyu LUO	
International Joint Conference on Artificial Intelligence (IJCAI 2025)	
<b>Boosting Text-to-Chart Retrieval through Training with Synthesized Semantic Insights</b>	<a href="#">Link ↗</a>
Yifan WU, Lutao YAN, Yizhang ZHU, Yinan MEI, Jiannan WANG, Nan TANG, Yuyu LUO	
Under Review	
<b>AskChart: Universal Chart Understanding through Textual Enhancement</b>	<a href="#">Link ↗</a>
Xudong YANG, Yifan WU*, Yizhang ZHU*, Nan TANG, Yuyu LUO	
Under Review	
<b>SRAG: Structured Retrieval-Augmented Generation for Multi-Entity Question Answering over Wikipedia Graph</b>	<a href="#">Link ↗</a>
Teng LIN, Yizhang ZHU, Yuyu LUO, Nan TANG	
Under Review	

## Goldman: Reading the Fed, Riding the Trend in Gold Markets with Multi-Agent LLMs [Link ↗](#)

Qiqi DUAN, Changlun LI, Yao SHI, Yizhang ZHU, Nan TANG, Yuyu LUO  
Under Review

## Concise Math Reasoning via Difficulty-Aware Distillation [Link ↗](#)

Yifan WU, Jingze SHI, Bingheng WU, Jiayi ZHANG, Xiaotian LIN, Yizhang ZHU, *et al.*, Yuyu LUO  
Under Review

## PROJECTS

### DeepEye - Core Developer [Link ↗](#)

*Visual Workflow Data Agent System (AI Agent 2025 Best Open-Source Project)*

- Developed a visual workflow data agent system to enable transparent and flexible data analysis.
- Designed a drag-and-drop workflow engine supporting zero-code editing and granular management, allowing users to visually build, debug, and monitor every step of the analysis pipeline in real-time.
- Implemented agentic workflow orchestration with multiple modes (ReAct, Planner-and-Executor, ReWOO, etc.), backed by a domain knowledge base for precise business logic alignment.

### GNN4SL - Project Leader [Link ↗](#)

*LLM-Enhanced Semantic-Aware Graph Learning for Schema Linking in NL2SQL*

- Reformulated schema linking as a graph link prediction problem, fusing LLM-generated semantic embeddings with GNN structural learning (GCN, GAT, RGAT) to align queries with schema elements.
- Constructed graph-structured datasets based on Spider and Bird benchmarks to validate the effectiveness and efficiency in capturing complex schema relationships.

## FUNDINGS AND AWARDS

AI Agent 2025 Competition - <i>Best Open-source Project Award</i>	<b>2025.10</b>
HKUST(GZ) Ph.D. Student Fellowship	<b>2025.09</b>
Data Science and Analytics Thrust Volunteer Grant	<b>2025.05</b>
Greater Bay Area CS Academic Poster Competition - <i>Most Popular Poster Award</i>	<b>2025.03</b>
HKUST(GZ) Red Bird M.Phil. Studentship	<b>2023.09</b>
Excellent Graduates of Chongqing University	<b>2023.06</b>
Comprehensive Scholarship of Chongqing University	<b>2022.09</b>
National Mathematical Contest of Modeling - <i>First Prize in Chongqing</i>	<b>2021.10</b>
National Undergraduate Innovation and Entrepreneurship Project - <i>¥50,000 funding</i>	<b>2021.05</b>

## ACADEMIC ACTIVITIES

### Talks & Seminars:

- Huawei Squirrel Webinar (Invited) - *A Survey of Data Agents: Emerging Paradigm or Overstated Hype?*
- HKUST(GZ) RBM Project Roadshow - *Towards Automated Statistical Analysis: The Role of LLMs*

### Teaching Assistant:

- DSAA6100: Data-centric AI Practical Lab (Master's Level), HKUST(GZ)

Academic Services: NeurIPS 2025 Reviewer

## SKILLS

English Proficiency: IELTS: 7.0 (Listening: 7.5, Reading: 8.5, Speaking: 6, Writing: 6.5)

### Professional Skills:

- Programming: Python, SQL, Java, C/C++, Verilog, JavaScript
- AI/Data Science: PyTorch, PEFT, vLLM, TRL, PyG, LangChain; Hadoop, Spark
- Development: Git; FastAPI; Vue, Streamlit; JMeter
- Computer Architecture/Hardware: FPGA, Vivado; Arduino

Academic Writing Skills: L<sup>A</sup>T<sub>E</sub>X, Microsoft Visio, OmniGraffle, Figma