

Yizhang ZHU

🏠 Homepage ✉️ yzhu305@connect.hkust-gz.edu.cn 🌐 Github 📄 Google Scholar

EDUCATION




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

EXPERIENCE






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

SELECTED PUBLICATIONS

Fist-authored Publications

- [1] A Survey of Data Agents: Emerging Paradigm or Overstated Hype? [Link](#) 
Yizhang ZHU, Liangwei WANG, Chenyu YANG, *et al.*, Guoliang LI, Yuyu LUO
Preprint (Survey)
- [2] EllieSQL: Cost-Efficient Text-to-SQL with Complexity-Aware Routing [Link](#) 
Yizhang ZHU, Runzhi JIANG, Boyan LI, Nan TANG, Yuyu LUO
Conference on Language Modeling (COLM 2025)
- [3] Are Large Language Models Good Statisticians? [Link](#) 
Yizhang ZHU, Shiyin DU, Boyan LI, Yuyu LUO, Nan TANG
Advances in Neural Information Processing Systems (NeurIPS 2024)

Co-authored Publications (View all in [Google Scholar](#))

- [4] LEAD: Iterative Data Selection for Efficient LLM Instruction Tuning [Link](#) 
Xiaotian LIN, Yanlin QI, Yizhang ZHU, Themis Palpanas, Chengliang CHAI, Nan TANG, Yuyu LUO
International Conference on Very Large Data Bases (VLDB 2026)
- [5] RAMer: Reconstruction-based Adversarial Model for Multi-party Multi-modal Multi-label Emotion Recognition [Link](#) 
Xudong YANG, Yizhang ZHU, Nan TANG, Yuyu LUO
International Joint Conference on Artificial Intelligence (IJCAI 2025)
- [6] DeepEye: A Steerable Self-driving Data Agent System [Link](#) 
Boyan LI, Yiran PENG, Yupeng XIE, Sirong LU, Yizhang ZHU, Xing MU, Xinyu LIU, Yuyu LUO
ACM SIGMOD International Conference on Management of Data (SIGMOD 2026 Demo)
- [7] OmniMoE: An Efficient MoE by Orchestrating Atomic Experts at Scale [Link](#) 
Jingze SHI, Zhangyang PENG, Yizhang ZHU, Yifan WU, Guang LIU, Yuyu LUO
Under Review
- [8] AskChart: Universal Chart Understanding through Textual Enhancement [Link](#) 
Xudong YANG, Yifan WU*, Yizhang ZHU*, Nan TANG, Yuyu LUO
Under Review

PROJECTS

DeepEye - Core Developer

[Link](#) 

A Steerable Self-driving Data Agent System (AI Agent 2025 Best Open-Source Project)

- A workflow-centric data agent system that orchestrates complex analytics as transparent DAGs, enabling joint analysis across heterogeneous sources such as databases, files, and knowledge bases.
- Features a database-inspired Workflow Engine (Compiler, Validator, Optimizer, Executor) that ensures structural correctness and efficiency through runtime topological optimization and parallel execution.
- Through hierarchical reasoning with context isolation and a Unified Node Protocol, synthesizes structured and unstructured data insights into Data Videos, interactive Dashboards, and Analytical Reports, moving beyond text-only analysis.

GNN4SL - Project Leader

[Link](#) 

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

- A semantic-aware schema linking approach that recasts NL2SQL schema linking as a graph link-prediction task, aligning user queries with database schema elements via a unified graph formulation.
- Combines LLM-derived semantic embeddings with GNN-based structural learning (GCN/GAT/RGAT) to capture both meaning-level similarity and complex relational dependencies among schema components.
- Builds graph-structured datasets and evaluation pipelines on Spider and BIRD to study effectiveness and efficiency under realistic multi-table schemas and challenging linking patterns.

FUNDINGS AND AWARDS

International Exhibition of Inventions Geneva - <i>Silver Medal</i>	2026.03
AI Agent 2025 Competition - <i>Best Open-source Project Award</i>	2025.10
HKUST(GZ) Ph.D. Student Fellowship	2025.09
Data Science and Analytics Thrust Volunteer Grant	2025.05
Greater Bay Area CS Academic Poster Competition - <i>Most Popular Poster Award</i>	2025.03
HKUST(GZ) Red Bird M.Phil. Studentship	2023.09
Excellent Graduates of Chongqing University	2023.06
Comprehensive Scholarship of Chongqing University	2022.09
National Mathematical Contest of Modeling - <i>First Prize in Chongqing</i>	2021.10
National Undergraduate Innovation and Entrepreneurship Project - <i>¥50,000 funding</i>	2021.05

ACADEMIC ACTIVITIES

Talks & Seminars:

- Huawei Squirrel Webinar (Invited) - *A Survey of Data Agents: Emerging Paradigm or Overstated Hype?*
- Huawei Paper Discussion Session: Insights into VLDB 2025 Industry Papers
- HKUST(GZ) RBM Project Roadshow - *Towards Automated Statistical Analysis: The Role of LLMs*

Teaching Assistant:

- DSAA6100: Data-centric Artificial Intelligence 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, DeepSpeed, PyG, LangChain; Hadoop, Spark
- Development: Git; FastAPI; Vue, Streamlit; JMeter
- Computer Architecture/Hardware: FPGA, Vivado; Arduino

Academic Writing Skills: L^AT_EX, Microsoft Visio, OmniGraffle, Figma