

Yizhang ZHU

☎ +86 19936076638 ✉ yzhu305@connect.hkust-gz.edu.cn 🏠 Homepage 🐙 Github









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	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

PUBLICATIONS

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

PROJECTS

EllieSQL

[Link](#)

Cost-Efficient Text-to-SQL with Complexity-Aware Routing

- Proposed a routing framework to optimize computational costs in Text-to-SQL by directing queries to suitable pipelines based on estimated complexity.
- Introduced Token Elasticity of Performance (TEP), a novel metric evaluating cost-efficiency by balancing performance gains and token usage.
- Investigated multiple router implementations, including classification-based (KNN, SFT), cascading, and preference learning-based (pairwise ranking, DPO) routers.
- Achieved > 40% reduction in token costs without compromising performance on Bird benchmark, improving TEP by 2× over non-routing approaches.

GNN4SL

[Link](#)

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

- Reformulated schema linking task as a link prediction problem in graph learning, where the objective was to establish connections between natural language query nodes and schema element nodes.
- Utilized large language models to generate semantic vector embeddings, thereby enhancing the representation of semantic information.
- Constructed a graph dataset based on the Spider and Bird training sets to train GNN models (GCN, GAT, and RGAT), enabling a more effective capture of schema structural information.

StatQA

[Link](#)

Benchmarking LLMs' Capabilities in Statistical Analysis

- StatQA Benchmark: Introduced a pipeline to synthesize a high-quality StatQA dataset, novelly curated for testing LLMs in specialized statistical analysis involving assessment of method applicability.
- Extensive Experiments: Systematically evaluated representative open-source and proprietary LLMs to establish our benchmark, also investigated the impact of in-context learning and supervised fine-tuning.
- Comparative Study between Humans and LLMs: Highlighted distinct strengths and weaknesses between humans and LLMs, revealed the potential for complementarity and collaboration.
- Explored and discussed research opportunities in this field.

FUNDINGS AND AWARDS

HKUST(GZ) Ph.D. Student Fellowship	2025.09 - Present
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 - 2025.06
Excellent Graduates of Chongqing University	2023.06
General 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

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^AT_EX, Microsoft Visio, OmniGraffle, Figma