

## YIZHANG ZHU

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

## EDUCATION

<b>The Hong Kong University of Science and Technology (Guangzhou)</b> <i>M.Phil. Student in Data Science and Analytics</i> <i>Supervisor: Prof. Yuyu LUO, Co-supervisor: Prof. Nan TANG</i>	<b>2023.9 - Present</b> <i>GPA 4.00/4.3</i>
<b>Chongqing University</b> <i>B.Eng. in Computer Science and Technology</i>	<b>2019.9 - 2023.6</b> <i>GPA 3.61/4.0</i>
<b>National University of Singapore Chongqing Research Institute</b> <i>Visiting Student in Computer Engineering Joint Program</i> <i>Supervisor: Prof. Yung Chii LIANG</i>	<b>2022.9 - 2023.5</b> <i>GPA 89.97/100</i>

## PUBLICATIONS

<b>Are Large Language Models Good Statisticians?</b> 🌐 <i>Yizhang ZHU, Shiyin DU, Boyan LI, Yuyu LUO, Nan TANG</i>	<b>NeurIPS'24</b>
<b>EllieSQL: Cost-Efficient Text-to-SQL with Complexity-Aware Routing</b> 🌐 <i>Yizhang ZHU, Runzhi JIANG, Boyan LI, Nan TANG, Yuyu LUO</i>	<b>Under Review</b>
<b>AskChart: Universal Chart Understanding through Textual Enhancement</b> <i>Xudong YANG, Yifan WU, Yizhang ZHU, Nan TANG, Yuyu LUO</i>	<b>Under Review</b>
<b>RAMer: Reconstruction-based Adversarial Model for Multi-party Multi-modal Multi-label Emotion Recognition</b> <i>Xudong YANG, Yizhang ZHU, Nan TANG, Yuyu LUO</i>	<b>Under Review</b>
<b>SRAG: Structured Retrieval-Augmented Generation for Multi-Entity Question Answering over Wikipedia Graph</b> <i>Teng LIN, Yizhang ZHU, Yuyu LUO, Nan TANG</i>	<b>Under Review</b>

## PROJECTS

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

- The schema linking task in NL2SQL was reformulated as a link prediction problem within the framework of graph learning, where the objective was to establish connections between NL query nodes and schema element nodes.
- The powerful natural language understanding capabilities of large language models were utilized to generate semantic vector embeddings, thereby enhancing the representation of semantic information.
- A graph dataset was constructed based on the Spider and Bird training sets to train a GNN model, enabling a more effective capture of schema structural information.

**StatQA: Benchmarking LLMs' Capabilities in Statistical Analysis Tasks**

- StatQA Benchmark: Introduced a pipeline to synthesize 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.
- Summarized key findings and discussed research opportunities in this field.

## FUNDINGS AND AWARDS

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 - 2025.06</b>
Excellent Graduates of Chongqing University	<b>2023.06</b>
General Scholarship of Chongqing University	<b>2022.09</b>
College Students Big Data Challenging Competition - <i>National Third Prize</i>	<b>2021.11</b>
National Mathematical Contest of Modeling - <i>First Prize in Chongqing</i>	<b>2021.10</b>
"Internet+" College Students Innovation Competition - <i>Silver Prize in Chongqing</i>	<b>2021.08</b>
National Undergraduate Innovation and Entrepreneurship Project - <i>¥50,000 funding from Chongqing University</i>	<b>2021.05</b>
China Collegiate Computing Contest - <i>Second Prize in Southwest Division</i>	<b>2021.08</b>