Junhao Zhu

Email: zhujunhao@zju.edu.cn | **Github**: github.com/JunHao-Zhu | **Homepage**: junhao-zhu.github.io

Education

Zhejiang University   2022/09 – 2027/06 (expected)
Ph.D. Student in Computer Science and Technology   Hangzhou, China
- Advisor: Lu Chen

Zhejiang University of Technology   2018/09 – 2022/06
B.E. in Automation Engineering   Hangzhou, China
- Graduated with the highest distinction.
- Advisor: Qi Xuan

Research Interests

Intelligence-Native Data Lake Management, Data Integration, Data Preparation, Data-centric AI

Publications

[1] **Junhao Zhu**, Lu Chen, Xiangyu Ke, Ziquan Fang, Tianyi Li, Yunjun Gao, Christian S. Jensen. "Beyond Relational: Semantic-Aware Multi-modal Analytics with LLM-Native Query Optimization". *In Proceedings of the ACM on Management of Data 4, 1* (**SIGMOD'26**): Article 14, 25 Pages.

[2] **Junhao Zhu**, Tao Wang, Lu Chen, Ziquan Fang, Yunjun Gao, Tianyi Li. "Toast: Task-Oriented Multi-dimensional Augmentation for Spatio-Temporal Trajectory Data". *In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2* (**SIGKDD'25**): 4180–4191.

[3] **Junhao Zhu**, Tao Wang, Danlei Hu, Ziquan Fang, Lu Chen, Yunjun Gao, Tianyi Li, Christian S. Jensen. "T-Assess: An Efficient Quality Assessment System Tailored for Trajectory Data". *In Proceedings of the 51th International Conference on Very Large Data Bases* (**VLDB'25**), 18(3): 666–674.

[4] **Junhao Zhu**, Yuren Mao, Lu Chen, Congcong Ge, Ziheng Wei, Yunjun Gao. "FusionQuery: On-demand Fusion Queries over Multi-source Heterogeneous Data". *In Proceedings of the 50th International Conference on Very Large Data Bases* (**VLDB'24**), 17(6): 1337–1349.

[5] Pengfei Wang, Xiaocan Zeng, Lu Chen, Fan Ye, Yuren Mao, **Junhao Zhu**, Yunjun Gao. "PromptEM: Prompt-tuning for Low-resource Generalized Entity Matching". *In Proceedings of the 49th International Conference on Very Large Data Bases* (**VLDB'23**), 16(2): 369–378.

[6] **Junhao Zhu**, Jinhuan Wang, Yalu Shan, Shanqing Yu, Guanrong Chen, Qi Xuan. "DeepInsight: Topology Changes Assisting Detection of Adversarial Samples on Graphs". *IEEE Transactions on Computational Social Systems* (**TCSS**), 11(1): 76–88, 2024.

Invited Talks

- "Nirvana: A LLM-powered Semantic Data Processing Computing Framework", **Alibaba**, Dec. 2025
- "An Efficient Data Quality Assessment System Tailored for Trajectory Data", **VLDB**, London, Sept. 2025
- "FusionQuery: On-demand Fusion Queries over Multi-source Heterogeneous Data", **VLDB**, Guangzhou, Aug. 2024
- "Conflict-resolved Queries over Multi-source Heterogeneous Data", **Huawei**, Jun. 2023
- "Weak Supervision for Entity Matching", **Huawei**, April 2023

Honors & Awards

- Provincial Outstanding Undergradates, Zhejiang University of Technology   2022
- Provincial Government Scholarship, Zhejiang University of Technology   2021 - 2022

Experience

Visiting student, *Tsinghua University*   2025/08 – current
Visiting student in database group (work with Guoliang Li)   Beijing, China
- Work on LLM-powered semantic data analytics processing.

Research intern, *Huawei*   2022/06 – 2022/09
Research Intern in Data Intelligence Innovation Lab (working with Ziheng Wei)   Hangzhou, China
- Customize a BERT-based language model on Huawei's data and support semantic table search over Huawei's data lake.

Skills

Programming Languages: Python, C/C++, Java
Tech Skills: PyTorch, Tensorflow, Linux, Git, Latex