

# ZICHUAN LIU

☺ Aspiring Data Scientist

👤 25/05/2000

✉ zichuanliu@mail.nju.edu.cn

🏠 zichuan-liu.github.io

🔍 google scholar

🐙 github.com/zichuan-liu

## 🎓 EDUCATION

### Nanjing University

Master of Electronic Engineering

09/2022 - 06/2025

- Grade Point Average (GPA): 89.38/100
- First-Class Scholarship in 2022 & 2023
- Supervised by Prof. Chunlin Chen

### Wuhan University of Technology

Bachelor of Software Engineering

09/2018 - 06/2022

- GPA: 91.10/100, Overall Ranking: 2/279, Grade Ranking: 15/279
- First-Class Scholarship, Top-up Scholarship, and Outstanding Student
- Supervised by A/Prof. Rui Zhang

## ♡ RESEARCH INTERESTS

My research interests focus on eXplainable Artificial Intelligence, Trustworthy LLM, and Reinforcement Learning, especially with applications in multiple sequences, offline models, and dynamic environments, allowing them to explicitly adapt to the non-stationary task distributions in real-world.

## 📖 PUBLICATIONS AND PREPRINTS

**Accepted Paper** (*ICML'24\*2, ICLR'24\*1, ICML'23\*1, etc.*, ♡ are representative papers.)

- ♡ **Zichuan Liu**, Tianchun Wang, Jimeng Shi, Xu Zheng, Zhuomin Chen, Lei Song, Wenqian Dong, Jayantha Obeysekera, Farhad Shirani, Dongsheng Luo. TimeX++: Learning Time-Series Explanations with Information Bottleneck. In *International Conference on Machine Learning*, Vienna, Austria, 2024. (*ICML'24, CCF-A*)
- Yifan Xia, Xianliang Yang, **Zichuan Liu**, Zhihao Liu, Lei Song, Jiang Bian. Rethinking Post-Hoc Search-Based Neural Approaches for Solving Large-Scale Traveling Salesman Problems. In *International Conference on Machine Learning*, Vienna, Austria, 2024. (*ICML'24 Oral, CCF-A, Top 2%*)
- ♡ **Zichuan Liu**, Yingying Zhang, Tianchun Wang, Zefan Wang, Dongsheng Luo, Mengnan Du, Min Wu, Yi Wang, Chunlin Chen, Lunting Fan, Qingsong Wen. Explaining Time Series via Contrastive and Locally Sparse Perturbations. In *International Conference on Learning Representations*, Vienna, Austria, 2024: 1-21. (*ICLR'24, CCF-A*)
- Linjie Xu, **Zichuan Liu**, Alexander Döckhorn, Diego Perez-Liebana, Jinyu Wang, Lei Song, Jiang Bian. Higher Replay Ratio Empowers Sample-Efficient Multi-Agent Reinforcement Learning. In *IEEE Conference on Games*, Milan, Italy, 2024. (*IEEE CoG'24*)
- ♡ **Zichuan Liu**, Yuanyang Zhu, Chunlin Chen. NA<sup>2</sup>Q: Neural Attention Additive Model for Interpretable Multi-Agent Q-Learning. In *International Conference on Machine Learning*, Honolulu, Hawaii, USA, 2023: 22539-22558. (*ICML'23, CCF-A*)
- **Zichuan Liu**, Rui Zhang, Chen Wang, Zhu Xiao, Hongbo Jiang. Spatial-Temporal Conv-sequence Learning with Accident Encoding for Traffic Flow Prediction. *IEEE Transactions on Network Science and Engineering*, 2022, 9(3): 1765-1775. (TNSE, SCI, JCR-Q1, IF=6.6)
- **Zichuan Liu**, Zhaoyang Wu, Meng Wang, Rui Zhang. Multi-View Spatial-Temporal Model for Travel Time Estimation. In *International Conference on Advances in Geographic Information Systems*, Beijing, China, 2021: 646-649. (*SIGSPATIAL'21, CCF-C*)

### Under Review

- Zhihao Liu, Xianliang Yang, **Zichuan Liu**, Yifan Xia, Wei Jiang, Yuanyu Zhang, Lijuan Li, Guoliang Fan, Lei Song, Jiang Bian. Knowing What Not to Do: Leverage Language Model Insights for Action Space Pruning in Multi-agent Reinforcement Learning. *arXiv preprint arXiv:2405.16854*, 2024.
- ♡ **Zichuan Liu**, Zefan Wang, Linjie Xu, Jinyu Wang, Lei Song, Tianchun Wang, Chunlin Chen, Wei Cheng, Jiang Bian. Protecting Your LLMs with Information Bottleneck. *arXiv preprint arXiv:2404.13968*, 2024.
- Zefan Wang, **Zichuan Liu**, Yingying Zhang, Aoxiao Zhong, Lunting Fan, Lingfei Wu, Qingsong Wen. RAgent: Cloud Root Cause Analysis by Autonomous Agents with Tool-Augmented Large Language Models. *arXiv preprint arXiv:2310.16340*, 2023.

- **Zichuan Liu**, Yuanyang Zhu, Zhi Wang, Yang Gao, Chunlin Chen. MIXRTs: Toward Interpretable Multi-Agent Reinforcement Learning via Mixing Recurrent Soft Decision Trees. *arXiv preprint* arXiv:2209.07225, 2022.
- Yuanyang Zhu, Qingpeng Zhao, **Zichuan Liu**, Zhi Wang, Chunlin Chen. Boosting Value Decomposition via Unit-Wise Attentive State Representation for Cooperative Multi-Agent Reinforcement Learning. *arXiv preprint* arXiv:2305.07182, 2023.

## INTERNSHIPS

**Microsoft**      Microsoft Research Asia      Research Intern      11/2023 - 05/2024

- I am currently as a research intern at MSRA Machine Learning Area, where I am advised by Dr. Lei Song and managed by Dr. Jiang Bian, also work close with Dr. Wei Cheng@NEC Labs.
- Two papers (explaining time series, rethink large-scale traveling salesman problems) were accepted by ICML'24, one of which I was the first author.
- One paper was accepted by IEEE COG'24 about sample efficient in RL.
- I completed a project to defense against jailbreaks of large language models, see the *preprint paper* here! Meanwhile, two papers (XAI for defending LLM, LLM guides RL) were submitted to NeurIPS'24.

**Alibaba Group**      DAMO & Alibaba Cloud      Research Intern      06/2023 - 11/2023

- I was a research intern in XAI for Time Series, advised by Dr. Qingsong Wen@DAMO Academy, also worked close with Dr. Tianchun Wang@PSU, A/Prof. Dongsheng Luo@FIU, A/Prof. Mengnan Du@NJIT.
- One paper was accepted by ICLR'24 on explaining time series predictions using contrastive and locally sparse perturbations (see *this slide* for more details).
- I had involved in writing a paper on tool-augmented LLMs for cloud root cause analysis, which has been submitted to CIKM'24.

**Wuhan Token Group**      Tech. Dept.      Android Engineer      03/2019 - 11/2019

- As a core member of the class schedule client, I was independently responsible for the development of seven pages. The daily activity number of its product reached 20000+.

## INVITED TALK

- “Perturbation-based Techniques for Explaining Sequence Predictions”, Invited **Talk @ Microsoft Research Asia**, April, 2024. [Slides]

## COMPETITIONS

 **Kaggle**: Kaggle Expert, Highest Ranking: 757th/213,123

- 2019 Data Science Bowl **TOP 1% & UP2ND**, DSB Student Award 3,000\$, 01/2020
- University of Liverpool - Ion Switching **TOP 1% & SOLO Silver**, 04/2020
- M5 Forecasting - Uncertainty **TOP 4% & Silver**, 07/2020

 **Data Mining**:

- ACM SIGSPATIAL 2021 GISCUPT **TOP 4th/1295**, Second Runner-up, Award 2,500\$, 08/2021
- China Collegiate Computing Contest **TOP 2nd/4133**, Award 10,000 ¥, 08/2020
- CCF Big Data & Computing Intelligence Contest **TOP 1% & 3rd/2703**, 12/2019

 **Mathematica Contest**:

- China Undergraduate Mathematical Contest in Modeling (CUMCM 2020) **National Second Prize**, 10/2020
- Mathematica Contest in Modeling (MCM 2020) **Honorable Mention**, 03/2020

## SERVICES

- Reviewer, Neural Information Processing Systems (NeurIPS), 2024.
- Reviewer, IEEE Transactions on Network Science and Engineering (TNSE), 2022.
- External reviewer, IEEE Transactions on Computational Social Systems (TCSS), 2022.

## MISCELLANEOUS

- Swimming is my primary form of exercise. I once achieved sixth place in the freestyle event at the university swimming competition in Hubei Province, and also successfully swam across the Yangtze River in 2019.