

# ZICHUAN LIU

☺ Aspiring Data Scientist

👤 25/05/2000

✉ zichuanliu@smail.nju.edu.cn

🏠 zichuan-liu.github.io

🔍 google scholar

🐙 github.com/zichuan-liu

## 🎓 EDUCATION

**Nanjing University** Master of Electronic and Information Engineering 09/2022 - 06/2025

- Grade Point Average (GPA): 89.375/100, IELTS Score: 7.0
- First-Class Scholarship, China National Scholarship (Top 0.2%)
- 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 Asso. Prof. Rui Zhang

## ♥ RESEARCH INTEREST

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

For your information, please check out my *Research Statement* and the slides of *Invited Talks* below.

## 📖 PUBLICATION AND PREPRINT

**Accepted Paper** (*TPAMI\*1, NeurIPS\*1, ICML\*3, ICLR\*2, CKIM\*1, etc.* ♡ are representative papers.)

- ♡ **[TPAMI'25]** Zichuan Liu, Yuanyang Zhu, Zhi Wang, Yang Gao, Chunlin Chen. MIXRTs: Toward Interpretable Multi-Agent Reinforcement Learning via Mixing Recurrent Soft Decision Trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.
- **[ICLR'25]** Tianchun Wang, Yuanzhou Chen, **Zichuan Liu**, Zhanwen Chen, Haifeng Chen, Xiang Zhang, Wei Cheng. Humanizing the Machine: Proxy Attacks to Mislead LLM Detectors. In *International Conference on Learning Representations*, 2025.
- ♡ **[NeurIPS'24]** **Zichuan Liu**, Zefan Wang, Linjie Xu, Jinyu Wang, Lei Song, Tianchun Wang, Chunlin Chen, Wei Cheng, Jiang Bian. Protecting Your LLMs with Information Bottleneck. In *Neural Information Processing Systems*, 2024.
- ♡ **[ICML'24]** **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*, 2024.
- **[ICML'24 Oral]** 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*, 2024.
- ♡ **[ICLR'24]** **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*, 2024.
- **[CIKM'24]** Zefan Wang, **Zichuan Liu**, Yingying Zhang, Aoxiao Zhong, Lunting Fan, Lingfei Wu, Qingsong Wen. RCagent: Cloud Root Cause Analysis by Autonomous Agents with Tool-Augmented Large Language Models. In *Conference on Information and Knowledge Management*, 2024.
- **[IEEE CoG'24]** Linjie Xu, **Zichuan Liu**, Alexander Dockhorn, Diego Perez-Liebana, Jinyu Wang, Lei Song, Jiang Bian. Higher Replay Ratio Empowers Sample-Efficient Multi-Agent Reinforcement Learning. In *IEEE Conference on Games*, 2024.
- ♡ **[ICML'23]** **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*, 2023.
- **[IEEE TNSE'22]** **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.
- **[SIGSPATIAL'21]** **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*, 2021.

## Under Review

- Tianchun Wang, **Zichuan Liu**, Yuanzhou Chen, Jonathan Light, Haifeng Chen, Xiang Zhang, Wei Cheng. Diversified Sampling Improves Scaling LLM inference. *arXiv preprint* arXiv:2502.11027, 2025.
- 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.
- Yuanyang Zhu, Qingpeng Zhao, **Zichuan Liu**, Zhi Wang, Chunlin Chen. Boosting Value Decomposition via Unit-Wise Attentive State Representation for Cooperative MARL. *arXiv preprint* arXiv: 2305.07182, 2023.

## EXPERIENCE

**Microsoft**      Microsoft Research Asia      Research Intern      10/2024 - 06/2025

- I joined the MSRA Industry Innovation Center in this winter-spring.
- I'm exploring why self-finetuning performs better than DPO or SimPO in real-world industrial applications – overfitting issues in preference optimization.

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

- Research intern at MSRA Machine Learning Area, working on XAI and RL for Large Language Models, advised by Dr. Lei Song and Dr. Jiang Bian.
- Milestones: ICML'24 Oral, ICML'24, NeurIPS'24, CoG'24, ArXiv'24, two of which I was first author.

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

- Research intern on XAI for Time Series, advised by Dr. Qingsong Wen.
- Milestones: ICLR'24, CIKM'24, one of which I was first author.

**IToken Group**      Technical Department      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. (including ICLR'24, ICML'24, NeurIPS'24) [Slides]
- “*Interpretable and Efficient Multi-Agent Reinforcement Learning*”, Invited Talk @ **Nanjing University**, February, 2024. (including ICML'23, ArXiv'22, CoG'24, ArXiv'24) [Slides]

## COMPETITION

 **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

 **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

 **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

## SERVICE

- Reviewer, International Conference on Machine Learning (ICML), Main Track and Position Track, 2025
- Reviewer, International Conference on Artificial Intelligence and Statistics (AISTATS), 2025
- Reviewer, International Conference on Learning Representations (ICLR), 2025
- Reviewer, Neural Information Processing Systems (NeurIPS), 2024
- Reviewer, IEEE Transactions on Network Science and Engineering (TNSE), 2022