Mr. Yuxiao Ye

184035@qq.com | +86 18519112663

PhD Student in Hong Kong University of Science and Technology

Research Interests: Deep Reinforcement Learning; LLM Post-training

🎓 EDUCATION

Hong Kong University of Science and Technology   **2025.08-Present**
PhD in Electronic and Computer Engineering   Supervisor: Prof. Ling Pan

Beijing Institute of Technology   (985 project university)   **2022.08-2025.06**
MSc in Computer Science and Technology   Supervisor: Prof. Chi (Harold) Liu, *FIET, FBCS*

Beijing Institute of Technology   **2018.08-2022.06**
BSc in Computer Science and Technology, Xuteli School (Honors College of BIT)   Average Score: 88.6/100 (rank: top 10%)

📄 PUBLICATIONS

Deep Reinforcement Learning

- **[CCF A - ICDE]** **Yuxiao Ye**, Chi Harold Liu, et al., "Exploring both Individuality and Cooperation for Air-Ground Spatial Crowdsourcing by Multi-Agent Deep Reinforcement Learning," in *IEEE ICDE*, 2023.
- **[CCF A - JSAC]** **Yuxiao Ye***, Hao Wang*, Chi Harold Liu, et al., "QoI-Aware Mobile Crowdsensing for Metaverse by Multi-Agent Deep Reinforcement Learning," in *IEEE Journal on Selected Areas in Communications (JSAC)*, 2024.
- **[CCF A - TMC]** **Yuxiao Ye***, Yuxuan Tian*, Chi Harold Liu, et al., "AoI-aware Air-Ground Mobile Crowdsensing by Multi-Agent Curriculum Learning with Collaborative Observation Augmentation," in *IEEE Transactions on Mobile Computing (TMC)*, 2025.
- **[CCF A - INFOCOM]** Zipeng Dai, Chi Harold Liu, **Yuxiao Ye**, et al., "AoI-minimal UAV Crowdsensing by Model-based Graph Convolutional Reinforcement Learning," in *IEEE INFOCOM*, 2022.

LLM Post-training

- **[EMNLP Findings]** Zhaohui Yang*, **Yuxiao Ye***, Shilei Jiang, Chen Hu, et al., "Unearthing Gems from Stones: Policy Optimization with Negative Sample Augmentation for LLM Reasoning," in *EMNLP 2025*.
- Haoran He*, **Yuxiao Ye***, Qingpeng Cai, Chen Hu, Binxing Jiao, Daxin Jiang, and Ling Pan. "Random Policy Valuation is Enough for LLM Reasoning with Verifiable Rewards." *submitted to ICLR 2025.*

Text-to-SQL

- **[ICLR Oral]** Fangyu Lei*, Jixuan Chen*, **Yuxiao Ye**, et al., "Spider 2.0: Can Language Models Resolve Real-world Enterprise Text-to-SQL Workflows?", in *ICLR 2025*.
- Bin Zhang*, **Yuxiao Ye***, et al., "SQLBench: A Comprehensive Evaluation for Text-to-SQL Capabilities of Large Language Models," *Arxiv Preprint*.
- Zhishuai Li*, Xiang Wang*, Jingjing Zhao*, Sun Yang*, Guoqing Du*, Xiaoru Hu*, Bin Zhang*, **Yuxiao Ye***, et al., "PET-SQL: A Prompt-enhanced Two-stage Text-to-SQL Framework with Cross-consistency," in *DASFAA 2025.*

🏆 HONORS and AWARD

Grand Prize in "China Collegiate Computing Contest - AI Innovation Contest" **(awarded 4/3400+)**   **2022**
National Scholarship (国家奖学金)   **2023 & 2024**
Outstanding Graduate of Beijing (北京市优秀毕业生)   **2025**

🧪 RESEARCH EXPERIENCES

Research Assistant, Mobile Crowdsensing and Combinatorial Optimization by (MA)DRL   **2022-2025**

- Proposed a MADRL framework, consisting of an intrinsic reward driven exploitation of individuality, enabling the accurate division of work, and a meta-learning based policy optimization, facilitating flexible agent's cooperation.
- Proposed a MADRL framework, with a traffic flow prediction mechanism based on spatial-temporal transformer, and a graph-based inter-agent communication method, to achieve efficient path planning for agents.
- Utilize transformer-based reinforcement learning to solve combinatorial optimization problems (particularly the two-echelon VRP), enhanced by a curriculum learning mechanism to mitigate non-stationarity among agents.

Intern, SenseTime, Large Language Model Group   **2023.12-2024.05**

- Constructed a new Text-to-SQL benchmark to mitigate overfitting in LLMs, conducted comprehensive evaluations on five Text-to-SQL sub-tasks across six LLMs, identified the distinct capabilities and limitations of LLMs, and proposed optimal in-context learning solutions tailored to each sub-task.

- Proposed an LLM-based Text-to-SQL framework, consisting of an enhancement of in-context learning and schema linking, and a cross-consistency mechanism across different models, which **achieves new SOTA results on the Spider benchmark with an accuracy of 87.6%.**

Intern, Stepfun   **2025**

- Current RL-based post-training methods for LLMs uniformly penalize all tokens in negative samples, despite the recognized value of certain steps like reflection and verification. We introduce a novel RL framework that **performs fine-grained segmentation and augmentation of these valuable steps within negative samples**, enhancing the reasoning capabilities of LLMs **(AIME24 +2%, AIME25 +4%)**.

- We developed **an offline RL approach with negative sample augmentation to train tool-calling LLM agents**. This method stimulates the agent to more actively utilize tools (e.g., search) for critical information retrieval, achieving performance gains while maintaining high sample efficiency **(GAIA +7%, XBench +8%).**

⚒ SKILLS

Programming: Python, C/C++/C#, Java, SQL, Matlab

Software: Pytorch, Tensorflow, Hugging Face Transformers, DeepSpeed

English Proficiency: IELTS 7.5