

乔丹

📞 178-0101-0541

✉ danqiao@link.cuhk.edu.cn

🏠 <https://qiaodan-cuhk.github.io>

🎓 教育经历

香港中文大学(深圳)	2022.09 – 2027.06
计算机科学与工程 博士在读 数据科学学院	深圳
• 主要研究方向: 多智能体强化学习, 合作机制设计, 大语言模型后训练;	
• 指导教师: 王昀翔, 查宏远	
北京航空航天大学	2014.09 – 2021.07
车辆工程、控制理论 本科 & 硕士 交通科学与工程学院	北京

🏢 实习经历

微软亚洲研究院, 算法实习生	2025.12 – 至今
• 研究主题: 通过数据选择提高 Agentic RL 训练效率。	
北京字节跳动科技有限公司, 算法实习生	2025.04 – 2025.11
• 研究主题: 多智能体强化学习微调与多大语言模型协同推理。	
深圳市固胜智能科技有限公司, 算法实习生	2024.12 – 2025.02
• 负责某赛车游戏的竞速策略强化学习分布式训练, 单车圈速策略达到专家级, 对标 Sony 相关研究。	
中国科学院自动化研究所, 研究助理	2021.09 – 2022.06
• 研究主题: 基于隐私保护的分布式网络化多智能体强化学习算法设计及收敛性分析。	

🔧 研究主题

大语言模型强化学习

- 基于 VERL 开发多智能体强化学习微调框架, 支持协调多个 LLM Agents 的端到端奖励驱动训练算法, 利用异构模型实现有效路径互补和合作推理能力提升。

扩散模型与离线强化学习

- 利用扩散模型的高表达能力捕捉多智能体互动数据的多模态行为策略, 通过联合策略正则分解避免离线强化学习的分布偏移问题, 在多个评估基准上实现显著性能提升。

合作多智能体强化学习的机制设计

- 基于社会科学的互惠理论 (Reciprocity Theory) 和声誉机制, 研究具有利益冲突的AI智能体如何学习社会规范并维持长期合作, 通过声誉信号实现陌生智能体之间的有效协同。

🏆 主要成果

Epistemic Gain, Aleatoric Cost: Uncertainty Decomposition in Multi-Agent Debate for Math Reasoning. D. Qiao, et, al. ICML Submission (Under Review), 2026. [arXiv:2603.01221](https://arxiv.org/abs/2603.01221).

Offline Multi-Agent Reinforcement Learning via Sequential Score Decomposition. D. Qiao, W. Li, S. Yang, H. Zha, B. Wang*. ICML Submission (Under Review), 2026. [arXiv:2505.05968](https://arxiv.org/abs/2505.05968).

STAR-MARL: LLM-based Sub-task Curricula Design. Zihao Li*, D. Qiao*, Arrasy Rahman, Stefanos Leonardos, Yali Du, Stefano V. Albrecht. [AAAI'26 LaMAS Workshop](https://arxiv.org/abs/2601.00000).

Multi-Agent Credit Assignment with Pretrained Language Models. W. Li, D. Qiao, B. Wang, X. Wang, B. Jin, H. Zha*. [\[AISTATS 2025\]](https://arxiv.org/abs/2505.05968).

Novel Saturated Nussbaum-type Function based Adaptive Distributed Consensus Control of Multi-agent Systems with Unknown Arbitrary Control Directions. D. Qiao, G. Wen, Z. Peng*. Preprint, 2021. [arXiv:2201.09453](https://arxiv.org/abs/2201.09453).

🔧 专业技能

- 编程语言与框架: VERL, PyTorch, Python
- 英语: IELTS 6.5