

# 张乔生

研究兴趣：强化学习与博弈、大模型推理与对齐、信息论与编码

## 工作经历

- |                           |                  |
|---------------------------|------------------|
| • 上海人工智能实验室，青年研究员-> 青年科学家 | 2022.5 至今        |
| • 新加坡国立大学，Research Fellow | 2019.9 - 2022.1  |
| • 佐治亚理工学院，Research Intern | 2018.6 - 2018.10 |

## 教育背景

- |                     |      |
|---------------------|------|
| • 博士学位：香港中文大学，信息工程系 | 2019 |
| • 学士学位：香港中文大学，信息工程系 | 2015 |

## 科研项目与荣誉

- |                                 |                   |
|---------------------------------|-------------------|
| • 优秀青年科学基金项目-海外（海外优青，经费 200 万元） | 2023.12 - 2026.12 |
| • 中国电子学会信息论分会青年新星奖（全国每年 1-2 位）  | 2024.11           |
| • 上海市海外高层次人才计划青年项目（白玉兰计划）       | 2023.7            |
| • 上海交通大学、复旦大学兼职博导               | 2024、2025         |

## 主要学术成果 (2023 至今)

- 多模态大模型（安全）推理能力增强
  - [SafeWork-R1: Coevolving Safety and Intelligence under the AI-45 Law](#)  
主要参与多模态大模型安全能力训练部分（通过大规模强化学习训练实现安全性能协同提升），并组织技术报告撰写
  - [MM-Eureka: Exploring the Frontiers of Multimodal Reasoning with Rule-based Reinforcement Learning](#)  
通过基于规则的强化学习，在开源社区较早实现了多模态大模型在多模态数学推理任务中的能力提升（2025.3 发布以来收获 1000+ Github Star, 谷歌学术 200+ 引用）
  - [CPGD: Toward Stable Rule-based Reinforcement Learning for Language Models](#)  
提出的强化学习算法 CPGD，显著提升大模型强化学习训练稳定性，并集成至 SafeWork-R1、Intern-S1 主线任务
- 人类反馈强化学习 (RLHF) 理论与算法
  - [Robust RLHF for Human Preference with Instance-Dependent Flipping](#)  
Y. Xu, X. Ye, Y. Chen, **Q. Zhang** *AAAI*, 2026
  - [Online Preference Alignment for Language Models via Count-based Exploration](#)  
C. Bai, Y. Zhang, S. Qiu, **Q. Zhang**, K. Xu, X. Li *ICLR*, 2025 (**Spotlight, Top 5.1%**)
  - [Sample-Efficient Reinforcement Learning from Human Feedback via Information-Directed Sampling](#)  
H. Qi, H. Yang, **Q. Zhang**, Z. Yang *IEEE Transactions on Information Theory*, 2025
- 强化学习与博弈
  - [Provably Efficient Information-Directed Sampling Algorithms for Multi-Agent Reinforcement Learning](#)  
**Q. Zhang**, C. Bai, S. Hu, Z. Wang, X. Li *Artificial Intelligence (AIJ)*, 2025
  - [On the Role of General Function Approximation in Offline Reinforcement Learning](#)  
C. Mao, **Q. Zhang**, Z. Wang, X. Li *ICLR*, 2024 (**Spotlight, Top 5%**)
  - [Constrained Ensemble Exploration for Unsupervised Skill Discovery](#)  
C. Bai, R. Yang, **Q. Zhang**, K. Xu, Y. Chen, T. Xiao, X. Li *ICML*, 2024
- 大模型多智能体、多模型路由
  - [The Avengers: A Simple Recipe for Uniting Smaller Language Models to Challenge Proprietary Giants](#)  
Y. Zhang, H. Li, C. Wang, L. Chen, Q. Zhang, et al. *AAAI*, 2026 (**Oral**)
  - [Do We Need So Many Samples? Multi-LLM Repeated Sampling Efficiently Scales Test-Time Compute](#)  
J. Chen, Z. Xun, B. Zhou, H. Qi, H. Zhang, Q. Zhang, et al. *AAAI*, 2026
  - [Multi-LLM-Agents Debate - Performance, Efficiency, and Scaling Challenges](#)  
H. Zhang, Z. Cui, **Q. Zhang**, S. Hu *ICLR Blogpost Track*, 2025

- 其他：聚类、图神经网络、信息论、信息安全等

- Graph Attention is Not Always Beneficial: A Theoretical Analysis of Graph Attention Mechanisms via Contextual-Stochastic Block Models  
Z. Ma, Q. Zhang, B. Zhou, Y. Zhang, S. Hu, Z. Wang *ICML*, 2025
- Exact Recovery in the General Hypergraph Stochastic Block Models  
Q. Zhang, V. Y. F. Tan *IEEE Transactions on Information Theory*, 2023
- Covert Communication with Mismatched Decoders  
Q. Zhang, V. Y. F. Tan *IEEE Transactions on Information Theory*, 2023