

牟倪 (mù ní)

📍 北京 · 清华大学 ✉ mn23@mails.tsinghua.edu.cn ☎ (+86) 182-5311-3690 📠 微信: MoonOutCloudBack
👤 个人主页 🎓 Google Scholar 🌐 GitHub 📝 技术博客

教育经历

- 博士研究生** 清华大学, 自动化系 2023.09 - 至今
- 18 岁攻读博士, 研究方向为 **大模型、强化学习**
 - 以第一 / 共同第一作者发表 **7 篇** CCF-A 会议和顶级国际期刊
- 本科** 东南大学, **少年班**, 计算机学院 2019.09 - 2023.06
- 14 岁考入东南大学少年班, 排名 3 / 275 (前 1.5%)

实习经历

- 实习生** 阿里巴巴 · **通义千问 Coding 团队** 2026.07 - 至今
- 面向 SWE Coding 任务, 开展 Agentic RL 数据生成、清洗与验证
 - 参与 SWE Coding 任务评测平台建设

科研成果

1. 强化学习方向: 一作 / 共同一作论文

- [1] CLARIFY: Contrastive Preference Reinforcement Learning for Untangling Ambiguous Queries** ICML 2025
Ni Mu*, Hao Hu*, Xiao Hu, Yiqin Yang, Bo Xu, Qing-Shan Jia

偏好强化学习 (Preference-based Reinforcement Learning, PbRL) 中, 人类难以比较行为相似的轨迹片段, 从而难以给出清晰偏好信号, 影响标签效率。针对这一问题, 我们在轨迹嵌入空间中设计对比学习 (Contrastive Learning) 方法, 选出人类可比较的片段对, 从而提升标签效率。

arxiv: 2506.00388 [🔗](#), Github [🔗](#)

- [2] STAIR: Addressing Stage Misalignment through Temporal-Aligned Preference Reinforcement Learning** NeurIPS 2025

Yao Luan*, **Ni Mu***, Yiqin Yang, Bo Xu, Qing-Shan Jia

我们发现, 在多阶段任务中, 偏好强化学习面临“阶段不对齐”问题: 比较来自不同阶段的行为, 既难以让人准确判断, 也难为策略优化提供有效信息。基于理论分析, 我们通过最小化片段对之间的时序距离 (Temporal Distance) 确保偏好比较在同一阶段, 从而提升学习效率。

arxiv: 2509.23802 [🔗](#), Github [🔗](#)

- [3] COLLIE: Guiding Skill Discovery in Semantically Coherent Latent Space** ICML 2026

Yao Luan*, **Ni Mu***, Hanfei Ge, Yiqin Yang, Bo Xu, Qing-Shan Jia

针对现有引导技能发现 (Skill Discovery) 方法依赖专家演示、在稀疏反馈下易过拟合的问题, 我们提出构建语义连贯的潜在空间, 并在其中实现免训练的引导信号构建, 基于极少人类标签即可学习安全、多样的技能。

OpenReview [🔗](#)

- [4] S-EPOA: Overcoming the Indistinguishability of Segments with Skill-Driven Preference-Based Reinforcement Learning** IJCAI 2025
Ni Mu*, Yao Luan*, Yiqin Yang, Bo Xu, Qing-Shan Jia
 针对偏好强化学习中，人类难以比较行为相似的轨迹片段的问题，我们设计无监督的技能发现 (Skill Discovery) 方法，选择“技能”差异较大的片段以供比较，从而提升人类标签效率。
 arxiv: 2408.12130 [🔗](#), Github [🔗](#)
- [5] Multi-Agent Reinforcement Learning via Agent-Specific Preference** IEEE TASE
Ni Mu*, Yao Luan*, Yiqin Yang, Qing-Shan Jia
 针对多智能体强化学习 (Multi-Agent RL) 中全局奖励函数难以设计的问题，我们构建 agent-wise 专家偏好反馈模型，并将其单调聚合为全局奖励函数，理论等价于学习 Nash 均衡策略。
 IEEE Transactions on Automation Science and Engineering, arxiv: 2608.08604 [🔗](#), Github [🔗](#)
- [6] Preference-based Multi-Objective Reinforcement Learning** IEEE TASE
Ni Mu*, Yao Luan*, Qing-Shan Jia
 针对多目标强化学习 (Multi-Objective RL) 中奖励函数难以设计的问题，我们提出了具有理论保证的多目标偏好学习范式：优化多目标偏好奖励模型，等价于求解策略空间的凸帕累托前沿。
 IEEE Transactions on Automation Science and Engineering, paper [🔗](#), arxiv: 2507.14066 [🔗](#)
- [7] MrCoM: A Meta-Regularized World-Model Generalizing Across Multi-Scenarios** AAAI 2026
 Xuantang Xiong*, **Ni Mu***, Runpeng Xie, Senhao Yang, Yaqing Wang, Lexiang Wang, Yao Luan, Siyuan Li, Shuang Xu, Yiqin Yang, Bo Xu
 针对基于模型的强化学习 (model-based RL) 中世界模型跨场景泛化难的问题，我们通过元状态正则化提取场景信息、元价值正则化对齐模型优化与多场景策略学习，构建跨场景世界模型。
 arxiv: 2511.06252 [🔗](#)
- [8] Safety-Guaranteed Policy Composition Via Generalized Policy Improvement for Autonomous Vehicles** CASE 2025 Outstanding WiRA Student Paper Award
Ni Mu, Yao Luan, Qing-Shan Jia
 我们建立了基于 Generalized Policy Improvement (GPI) 的策略集成方法，该方法具有安全性保障和性能保障，可以解决自动驾驶中“性能-安全性-样本效率”的“不可能三角形”。
 IEEE 21st International Conference on Automation Science and Engineering, paper [🔗](#)

2. 大模型方向：一作 / 共同一作论文

- [9] SunTzu: A Long-Horizon Decision Making Benchmark for Large Language Models** Preprint, Submitted to NeurIPS 2026
 Pengbo Shen*, Yaqing Wang*, **Ni Mu***, Yao Luan, Runpeng Xie, Senhao Yang, Lexiang Wang, Hao Hu, Shuang Xu, Yiqin Yang, Bo Xu
 为评测大模型 (LLM) 的长时 (long-horizon) 决策能力，我们提出了 SunTzu，一个专为评测大模型在星际争霸 II 游戏中表现的基准；进而设计一个分层决策 baseline Agent 框架，基于专家知识自我纠正、利用监督微调 (SFT) 自我提升，显著提升 LLM 的战略推理与执行能力。
 arxiv: 2508.10428 [🔗](#), Github [🔗](#)
- [10] Covering the Pareto Frontier with LLM-Coordinated Interpretable Policy Library** IFAC 2026
Ni Mu*, Yao Luan*, Qing-Shan Jia
 23rd IFAC World Congress
- [11] Learning Interpretable Index Policies via LLM-Driven Zero-Order Optimization** CASE 2026
 Yao Luan*, **Ni Mu***, Qing-Shan Jia

[12] LLM-Empowered Knowledge Graph Construction for Optimization Formalization in Smart Manufacturing CAC 2025

Ni Mu, Zeyuan Liu, Yuhang Zhu, Qing-Shan Jia

2025 China Automation Congress (CAC), paper [🔗](#)

3. 非一作工作，以及强化学习的工业应用

[13] GlobeDiff: State Diffusion Process for Partial Observability in Multi-Agent System ICLR 2026

Yiqin Yang*, Xu Yang*, Yuhua Jiang, **Ni Mu**, Hao Hu, Runpeng Xie, Ziyu Zhang, Siyuan Li, Yuan-Hua Ni, Qianchuan Zhao, Bo Xu

OpenReview [🔗](#)

[14] DAIL: Beyond Task Ambiguity for Language-Conditioned Reinforcement Learning NeurIPS 2025

Runpeng Xie*, Quanwei Wang*, Hao Hu, Zherui Zhou, **Ni Mu**, Xiyun Li, Yiqin Yang, Shuang Xu, Qianchuan Zhao, Bo Xu

arxiv: 2510.19562 [🔗](#), Github [🔗](#)

[15] A Transformer-based Thermal Surrogate Model for Cooling Control in Data Centers IEEE RA-L

Hanchen Zhou, **Ni Mu**, Qing-Shan Jia

IEEE Robotics and Automation Letters, paper [🔗](#)

[16] E-MAPP: Efficient Multi-Agent Reinforcement Learning with Parallel Program Guidance NeurIPS 2022

Can Chang, **Ni Mu**, Jiajun Wu, Ling Pan, Huazhe Xu

arxiv: 2212.02064 [🔗](#), Website [🔗](#), Github [🔗](#)

[17] MAGPIE: Utilizing Agent-Specific Preferences for Multi-Agent Reinforcement Learning CAC 2025

Ni Mu*, Yao Luan*, Qing-Shan Jia

2025 China Automation Congress (CAC), paper [🔗](#)

[18] Preference-Based Multi-Objective Reinforcement Learning with Explicit Reward Modeling CAC 2024

Ni Mu, Yao Luan, Qing-Shan Jia

2024 China Automation Congress (CAC), paper [🔗](#)

[19] Large-scale data center cooling control via sample-efficient reinforcement learning CASE 2024

Ni Mu, Xiao Hu, Qing-Shan Jia, Xun Zhu, Xiao He

IEEE 20th International Conference on Automation Science and Engineering, paper [🔗](#)

[20] Integrating mechanism and data: Reinforcement learning based on multi-fidelity model for data center cooling control CAC 2023

Ni Mu, Xiao Hu, Qing-Shan Jia

2023 China Automation Congress (CAC), paper [🔗](#)

[21] Addressing Coupling in Restless Multi-Armed Bandits by Finetuning Whittle Index CASE 2025

Yao Luan, **Ni Mu**, Qing-Shan Jia

IEEE 21st International Conference on Automation Science and Engineering, paper [🔗](#)

荣誉奖项

Outstanding WiRA Student Paper Award - IEEE CASE 2025	2025.08
清华大学综合奖学金 · 一等奖	2025.11
清华大学综合奖学金 · 一等奖	2024.12
东南大学最具影响力本科生	2023.05
东南大学校长奖学金	2021.12
国家奖学金（本科阶段）	2020.12