

# 李克钰

✉ chlorophyll@sjtu.edu.cn

✉ @chlorophtllwzh

☎ 13370856505

🌐 keyuli.space

☎ +86 13370856505

📖 4183134930

🔍 Google Scholar

🏠 Github

🤖 Huggingface



## 个人优势

- 以第一作者身份发表两篇 CCF A 类会议在 ACL 2026, ICLR 2026
- Claude Code, Codex 深度使用者, 熟悉当前最先进 AI Agent 的能力边界
- 熟练掌握 linux, pytorch, verl, llama-factory, slime 等框架, 对新事物学习能力强, 上手快

## 教育经历

- 2024 - 2029 **二年级博士研究生, 上海交通大学 & 上海创智学院 计算机科学与技术.**  
研究方向: AI 智能体, 多智能体系统, 大语言模型后训练与评测.  
导师: 王德泉老师, 刘鹏飞老师.
- 2020 - 2024 **本科生, 上海交通大学 数学与应用数学.**  
上海市优秀毕业生, 上海交大郭晨晨奖学金, 本科生奖学金, 优秀团员, 三好学生.  
学积分 92.6/100, GPA 4.03/4.3, 排名 4/45.  
毕业论文: 贝叶斯流网络的理解和可视化研究.
- 2017 - 2020 **山东省青岛第二中学**  
青岛二中优秀毕业生, 三好学生, 优秀团员.

## 实习经历

- 2025.5 - 现在 **LLM Agent 算法研究, 上海创智学院**  
上海交通大学/上海创智学院刘鹏飞老师, 生成式人工智能实验室 (GAIR Lab)  
研究大模型智能体后训练与评测, 代表作 Agencybench, LIM1, DatasetResearch
- 2024.6 - 2024.9 **量化研究员, 东吴证券金融科技量化研究部**  
利用深度学习模型挖掘因子, 使用 CNN, RNN, Transformer, KAN 等搭建模型,  
在 Level-2 实盘数据上进行训练, 预测 IC 相关系数, 给实盘交易提供信号

## 研究成果

### Conference Proceedings

- 1 K. Li, J. Gao, and D. Wang, "Aligned agents, biased swarm: Measuring bias amplification in multi-agent systems," in *The Fourteenth International Conference on Learning Representations*.
- 2 J. Gao, J. Zhao, K. Li, and D. Wang, "Kan-mixer: Kolmogorov-arnold networks for gene expression prediction in plant species," in *European Conference on Computer Vision*, Springer, 2024, pp. 135-150.

## Articles

- 1 M. Jiang, D. Fu, J. Shi, J. Zeng, W. Si, **K. Li**, X. Li, Y. Xiao, W. Li, D. Wang, et al., “Davinci-agency: Unlocking long-horizon agency data-efficiently,” *arXiv preprint arXiv:2602.02619*, 2026.
- 2 **K. Li**, J. Shi, Y. Xiao, M. Jiang, J. Sun, Y. Wu, S. Xia, X. Cai, T. Xu, W. Si, et al., “Agencybench: Benchmarking the frontiers of autonomous agents in 1m-token real-world contexts,” *arXiv preprint arXiv:2601.11044*, 2026.
- 3 **K. Li**, M. Jiang, D. Fu, Y. Wu, X. Hu, D. Wang, and P. Liu, “Datasetresearch: Benchmarking agent systems for demand-driven dataset discovery,” *arXiv preprint arXiv:2508.06960*, 2025.
- 4 Y. Wu, D. Fu, W. Si, Z. Huang, M. Jiang, **K. Li**, S. Xia, J. Sun, T. Xu, X. Hu, et al., “Innovatorbench: Evaluating agents’ ability to conduct innovative llm research,” *arXiv preprint arXiv:2510.27598*, 2025.
- 5 Y. Xiao, M. Jiang, J. Sun, **K. Li**, J. Lin, Y. Zhuang, J. Zeng, S. Xia, Q. Hua, X. Li, et al., “Limi: Less is more for agency,” *arXiv preprint arXiv:2509.17567*, 2025.

## 项目经历

- [ICLR 2026 第一作者] **Aligned Agents, Biased Swarm: Measuring Bias Amplification in Multi-Agent Systems**  
提出多智能体偏见评估基准 Discrim-Eval-Open，揭示复杂协作结构会放大而非缓解系统性偏见，为伦理鲁棒性评估奠定基线。
- [ACL 2026 第一作者] **Agencybench: Benchmarking the frontiers of autonomous agents in 1m-token real-world contexts**  
构建包含真实长周期任务的自动化评估基准 AgencyBench，系统揭示了不同大模型在资源效率、自我修正及框架适配上的差异。
- [ICLR 2026] **InnovatorBench: Evaluating Agents’ Ability to Conduct Innovative LLM Research**  
开发首个大模型端到端自动化科研评估平台 InnovatorBench，揭示了前沿模型在长周期科研规划与复杂算法探索中的核心局限。
- [Submitted to NeurIPS 2026 第一作者] **Datasetresearch: Benchmarking agent systems for demand-driven dataset discovery**  
提出首个评估智能体自主发现与合成数据集能力的基准 DatasetResearch，确立了严格基线并指明深层数据检索系统的发展方向。
- [Submitted to NeurIPS 2026 共同第一作者] **Limi: Less is more for agency**  
提出“少即是多 (LIMI)”训练范式，证明仅需极少量精选高质量数据即可激发卓越的智能体自主能力，打破传统数据扩展定律。
- [Submitted to ICML 2026] **daVinci-Agency: Unlocking Long-Horizon Agency Data-Efficiently**  
提出基于真实代码 PR 演进序列的数据合成框架 daVinci-Agency，有效提取长依赖监督信号，显著提升模型长视距规划与执行性能。