

# Keyu Li 李克钰

✉ chlorophyll@sjtu.edu.cn

✉ @chlorophtllwzh

☎ 13370856505

🌐 keyuli.space

☎ +86 13370856505

📖 4183134930

🔍 Google Scholar

🏠 Github

🤖 Huggingface



## Personal Strengths

- Published **two CCF A-tier conference papers** as the **first author** at ACL 2026 and ICLR 2026.
- Advanced user of Claude Code and Codex; highly familiar with the capability boundaries of AI Agents.
- Proficient in Linux, PyTorch, verl, LLaMA-Factory, and slime frameworks.

## Education

- 2024 – 2029 **2nd-Year Ph.D. Student, Shanghai Jiao Tong University & Shanghai Innovation Institute** Computer Science and Technology.  
Research Interests: *AI Agents, Multi-Agent Systems, LLM Post-Training and Evaluation.*  
Advisors: Prof. Dequan Wang, Prof. Pengfei Liu.
- 2020 – 2024 **B.S., Shanghai Jiao Tong University** Mathematics and Applied Mathematics.  
Shanghai Outstanding Graduate, SJTU Undergraduate Scholarship,  
Academic Score: 92.6/100, GPA: 4.03/4.3, Rank: 4/45.  
Bachelor's Thesis: *Understanding and Visualization of Bayesian Flow Networks.*
- 2017 – 2020 **Qingdao No. 2 High School, Shandong Province**  
Outstanding Graduate of Qingdao No. 2 High School

## Internship Experience

- 2025.05 – Present **Algorithm Researcher (LLM Agents), Shanghai Innovation Institute**  
Generative AI Research Lab (GAIR Lab). Advisor: Prof. Pengfei Liu.  
Researching post-training and evaluation of LLM agents. Representative works: AgencyBench, LIML, DatasetResearch.
- 2024.06 – 2024.09 **Quantitative Researcher, Fintech Department, Soochow Securities**  
Utilized deep learning models for factor mining; developed architectures including CNN, RNN, Transformer, and KAN. Trained models on Level-2 live trading data to predict Information Coefficient (IC) and provide signals for real-market execution.

## Research Achievements

### 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.*

## 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.

## Project & Publication Experience

- **[ICLR 2026, First Author] Aligned Agents, Biased Swarm: Measuring Bias Amplification in Multi-Agent Systems**  
Proposed Discrim-Eval-Open, a bias evaluation benchmark revealing that complex collaborative structures amplify rather than mitigate systemic biases in multi-agent systems.
- **[ACL 2026, First Author] AgencyBench: Benchmarking the frontiers of autonomous agents in 1M-token real-world contexts**  
Constructed AgencyBench to evaluate real-world, long-horizon tasks, revealing how large language models differ in resource efficiency, self-correction, and framework adaptation.
- **[ICLR 2026] InnovatorBench: Evaluating Agents’ Ability to Conduct Innovative LLM Research**  
Developed InnovatorBench, an automated research evaluation platform that reveals the core limitations of frontier LLMs in long-horizon planning and algorithmic exploration.
- **[Submitted to NeurIPS 2026, First Author] DatasetResearch: Benchmarking agent systems for demand-driven dataset discovery**  
Proposed DatasetResearch to evaluate agents’ abilities to autonomously discover and synthesize datasets, establishing strict baselines for future deep data retrieval systems.
- **[Submitted to NeurIPS 2026, Co-First Author] LIMi: Less is more for agency**  
Proposed the “Less is More” (LIMI) training paradigm, demonstrating that a minimal amount of curated, high-quality data can unlock exceptional agent autonomy and break traditional scaling laws.
- **[Submitted to ICML 2026] daVinci-Agency: Unlocking Long-Horizon Agency Data-Efficiently**  
Proposed daVinci-Agency, a data synthesis framework based on real-world Pull Request sequences that extracts long-dependency signals to enhance models’ long-horizon planning and execution.