Eric Hanchen Jiang

+1-626-629-6812 | ericjiang0318@ucla.edu | Homepage | Google Scholar

RESEARCH AREA

My research focuses on advancing large language models in reasoning, safety, and agentic capabilities, with the goal of building systems that are reliable, effective, and scalable for real-world applications. I am currently developing agent harnesses for tackling challenging mathematical problems, including open problems. Looking ahead, I aim to design efficient agent infrastructure that improves how LLM-based agents use tools, memory, and long-horizon reasoning.

EDUCATION

University of California, Los Angeles (UCLA) Aug. 2025 – Jun. 2030 (Expected)

- Ph.D. in Statistics and Data Science
- Advisors: Prof. Ying Nian Wu and Prof. Kai-Wei Chang

University of California, Los Angeles (UCLA) Jun. 2022 – Jun. 2025

- B.S. in Statistics & Data Science; B.S. in Mathematics of Computation | GPA: 3.85/4.00

SELECTED PUBLICATIONS

[1] **E.H. Jiang**, G. Wan, S. Yin, M. Li, Y. Wu, X. Liang, X. Li, Y. Sun, W. Wang, K.-W. Chang, Y.N. Wu. "Dynamic Generation of Multi-LLM Agents Communication Topologies with Graph Diffusion Models." **ACL Main**, 2026.

[2] **E.H. Jiang**, W. Ou, R. Liu, S. Pang, G. Wan, R. Duan, W. Dong, K.-W. Chang, X. Wang, Y.N. Wu, X. Li. "Mitigating Over-Refusal in Aligned Large Language Models via Inference-Time Activation Energy." **ACL Main**, 2026.

[3] **E.H. Jiang**, L. Li, R. Sun, X. Liang, Y. Li, Y. Wu, H. Luo, H. Li, Z. Zhang, Z. Kang, et al. "Agent Q-Mix: Selecting the Right Action for LLM Multi-Agent Systems through Reinforcement Learning." *Preprint*, 2026.

[4] **E.H. Jiang**, H. Luo, S. Pang, X. Li, Z. Qi, H. Li, C.-F. Yang, Z. Lin, X. Li, H. Xu, K.-W. Chang, Y.N. Wu. "Learning to Rank Chain-of-Thought: Using a Small Model." *Preprint*, 2026.

[5] X. Liang, Z.Z. Li, Z. Lin, **E.H. Jiang**, H. Zhang, Y. Shen, K.-W. Chang, Y.N. Wu, et al. "Training LLMs for Divide-and-Conquer Reasoning Elevates Test-Time Scalability." **ACL Main**, 2026.

[6] G. Wan, M. Zhou, Z. Wang, X. Shang, **E.H. Jiang**, G. Zhang, J. Bi, Y. Ma, et al. "DAWN: Distributed LLM Multi-Agent Workflow Synthesis." **AAAI**, 2026.

[7] X. Li, X. Chen, J. Fan, **E.H. Jiang**, M. Gao. "Multi-head Reward Aggregation Guided by Entropy." **AAAI**, 2026.

[8] G. Wan, L. Fu, H. Liu, Y. Jin, H.Y. Leong, **E.H. Jiang**, H. Geng, J. Bi, Y. Ma, et al. "Beyond Magic Words: Sharpness-Aware Prompt Evolving for Robust Large Language Models with TARE." **ICLR**, 2026.

[9] S. Chen, X. Li, M. Zhang, **E.H. Jiang**, Q. Zeng, C.-H. Yu. "CARES: Comprehensive Evaluation of Safety and Adversarial Robustness in Medical LLMs." **NeurIPS** Datasets & Benchmarks, 2025.

[10] Z. Zhang, C. Chow, Y. Zhang, Y. Sun, H. Zhang, **E.H. Jiang**, H. Liu, F. Huang, Y. Cui, O.H.M. Padilla. "Statistical Guarantees for Lifelong Reinforcement Learning Using PAC-Bayesian Theory." **AISTATS**, 2025.

[11] **E.H. Jiang**, A. Lizarraga. "SDSRA: A Skill-Driven Skill-Recombination Algorithm for Efficient Policy Learning." **ICLR** Tiny Papers, 2024.

WORK EXPERIENCE

Alibaba Cloud Group Jun. 2025 – Aug. 2025

Machine Learning Research Intern Mentor: Ranjie Duan

- Developed methods to mitigate jailbreaking attacks on Large Language Models.
- Built Agentic LLMs for personalized multi-turn dialogue systems.

AWARDS & HONORS

UCLA Graduate Dean's Scholar Award (GDSA) 2025
Princeton University Graduate Student Fellowship (Declined) 2025
Dean's Honors List, UCLA 2023 – 2024

PROFESSIONAL SERVICE

Conference Reviewer: AAAI 2024–2025, NeurIPS 2025, ICLR 2026, ICML 2026

Journal Reviewer: Transactions on Machine Learning Research (TMLR), STAT, Electronic Journal of Statistics