Jianliang He

Kline Tower, 219 Prospect Street, New Haven, CT 06511
Phone: (+1) 203-410-5714 | Email: jianliang.he@yale.edu | Website: jlianghe.github.io

RESEARCH INTERESTS

Machine Learning Theory; Mechanistic Interpretability; Large Language Model.

EDUCATION

Yale University 2024.9 - Present
Department of Statistics and Data Science New Haven, CT
Ph.D. in Statistics.
Advisor: Prof. Zhuoran Yang.

Fudan University 2020.9 - 2024.6
Department of Statistics and Data Science, School of Management Shanghai, China
B.S. in Statistics

RESEARCH PAPERS

* stands for equal contribution or alphabetical ordering.

1. **He, J.**, Wang, L., Chen, S., Yang, Z. “On the Mechanism and Dynamics of Modular Addition: Fourier Features, Lottery Ticket, and Grokking”. arXiv.2602.16849. Submitted, 2026.
2. Wei, J., Chen, S., **He, J.**, Yang, Z. “How Transformers Learn Causal Structures In-Context: Explainable Mechanism Meets Theoretical Guarantee”. *International Conference on Learning Representations (ICLR)*, 2026.
3. **He, J.**, Pan, X., Chen, S., Yang, Z. “In-Context Linear Regression Demystified: Training Dynamics and Mechanistic Interpretability of Multi-Head Softmax Attention”. arXiv.2503.12734. *International Conference on Machine Learning (ICML)*, 2025.
4. Qin, S.*, **He, J.***, Kuang, Q*., Gang, B, Xia, Y. “Data-light Uncertainty Set Merging with Admissibility”. arXiv.2410.12201. Submitted, 2024.
5. **He, J.***, Chen, S.*, Zhang, F., Yang, Z. “From Words to Actions: Unveiling the Theoretical Underpinnings of LLM-Driven Autonomous Systems”. arXiv.2405.19883. *International Conference on Machine Learning (ICML)*, 2024.
6. **He, J.**, Zhong, H., Yang, Z. “Sample-Efficient Learning of Infinite-Horizon Average-Reward MDPs with General Function Approximation”. arXiv.2404.12648. *International Conference on Learning Representations (ICLR)*, 2024.
7. Banerjeea, T.*, Gang, B.*, **He, J.***. “Harnessing the Collective Wisdom: Fusion Learning using Decision Sequences from Diverse Sources”. arXiv.2308.11026. *Biometrika*, 2026.

INDUSTRIAL EXPERIENCE

Machine Learning Engineer, Cisco Foundation AI Team, San Francisco 2025.6 - Present

- Leveraged post-training pipelines to build a reasoning Large Language Model (LLM) for cybersecurity domain. The technical report is available at arXiv.2601.21051.

TEACHING

Teaching Assistant, Yale University

- S&DS 241 Probability Theory Fall, 2025

- S&DS 265 Introductory Machine Learning Spring, 2026

Teaching Assistant, Fudan University

- MANA130083.01 Nonparametric Statistics Spring, 2023

SERVICE

Conference Reviewer: NeurIPS (2024-2025), ICLR (2025-2026), ICML (2025-2026).
Journal Reviewer: Management Science.