

Jianliang He

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

RESEARCH INTERESTS

Machine Learning Theory; Mechanistic Interpretability; Selective Inference.

EDUCATION

Yale University Department of Statistics and Data Science Ph.D. in Statistics	2024.9 - Present New Haven, CT
Fudan University Department of Statistics and Data Science, School of Management B.S. in Statistics	2020.9 - 2024.6 Shanghai, China

RESEARCH PAPERS

* stands for equal contribution or alphabetical ordering.

- He, J.**, Wang, L., Chen, S., Yang, Z. “On the Mechanism and Dynamics of Modular Addition: Fourier Features, Lottery Ticket, and Grokking”. Submitted, 2025.
- He, J.**, Pan, X., Chen, S., Yang, Z. “In-Context Linear Regression Demystified: Training Dynamics and Mechanistic Interpretability of Multi-Head Softmax Attention”. *International Conference on Machine Learning (ICML)*, 2025.
- Qin, S.*, **He, J.***, Kuang, Q., Gang, B., Xia, Y. “SAT: Data-light Uncertainty Set Merging via Synthetics, Aggregation, and Test Inversion”. Submitted to *Journal of the American Statistical Association*, 2024.
- He, J.***, Chen, S.*, Zhang, F., Yang, Z. “From Words to Actions: Unveiling the Theoretical Underpinnings of LLM-Driven Autonomous Systems”. *International Conference on Machine Learning (ICML)*, 2024.
- He, J.**, Zhong, H., Yang, Z. “Sample-Efficient Learning of Infinite-Horizon Average-Reward MDPs with General Function Approximation”. *International Conference on Learning Representations (ICLR)*, 2024.
- Banerjee, T.*, Gang, B.*, **He, J.***. “Harnessing the Collective Wisdom: Fusion Learning using Decision Sequences from Diverse Sources”. Under Revision at *Biometrika*, 2023.

WORK EXPERIENCE

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

TEACHING EXPERIENCE

Teaching Assistant, S&DS 2410 Probability Theory, Yale University	Fall, 2025
Teaching Assistant, MANA130083.01 Nonparametric Statistics, Fudan University	Spring, 2023

SERVICE

Conference Reviewer: NeurIPS (2024, 2025), ICLR 2025, ICML 2025.

Journal Reviewer: Management Science.