

Jianliang He

Kline Tower, 219 Prospect Street, New Haven, CT 06511

Phone: (+1) 203-410-5714 | Email: jianliang.he@yale.edu

RESEARCH INTERESTS

Mechanistic Interpretability; Machine Learning Theory; Selective Inference.

EDUCATION

Yale University	2024.9 - Present
Department of Statistics and Data Science	New Haven, CT
Ph.D. in Statistics	
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.

- 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”. Submitted to *Biometrika*, 2023.

HONORS & AWARDS

First Prize Scholarship for Outstanding Students (Top 1%), Fudan University	2022 - 2023
Outstanding Award, Fudan Undergraduate Research Opportunity Program (FDUROP)	2022, 2023

TEACHING EXPERIENCE

Teaching Assistant, MANA130083.01 Nonparametric Statistics, Fudan University	Spring, 2023
--	--------------

SKILLS & INTERESTS

Programming: Python, R, Matlab, SQL, C/C++, ~~Java~~ \LaTeX .
Languages: Chinese (Native), English (Fluent).

SERVICE

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