Wenbo Zhang

Department of Statistics, University of California Irvine
wenbz13@uci.edu Homepage

RESEARCH INTERESTS

My current research focuses on Reinforcement Learning from Human Feedback (RLHF) for aligning large language models (LLMs), with an emphasis on reward modeling. I am also deeply interested in trustworthy machine learning, particularly in areas such as uncertainty quantification (confidence calibration) and causal inference.

EDUCATION

University of California, Irvine *2021.9 - Present*
PhD Candidate, Statistics
Advisor: Prof. Hengrui Cai

University of Washington, Seattle *2019.9 - 2021.3*
Master of Science, Biostatistics

Xi'an Jiaotong-Liverpool University *2015.9 - 2019.7*
Bachelor of Science, Applied Mathematics

SELECTED PUBLICATIONS AND MANUSCRIPTS

* denotes equal contribution

LLMs & NLP

[1] Wenbo Zhang, Wenzhuo Zhou, Hengrui Cai, and Zhenglin Qi. (2025) "Towards Bridging the Gap Between Offline and Iterative Alignment via Preference Distillation." Under review in NeurIPS 2025.

[2] Wenbo Zhang, Hengrui Cai, and Wenyu Chen. (2025) "Beyond the Singular: The Essential Role of Multiple Generations in Effective Benchmark Evaluation and Analysis." Under review in ACL Rolling Review. Available in arXiv. [Link]

[3] Wenzhuo Zhou*, Wenbo Zhang*, and Hengrui Cai. (2024) "Regularized Offline Alignment for Language Models."

[4] Wenbo Zhang*, Zihang Xu*, and Hengrui Cai. (2024) "Defining Boundaries: A Spectrum of Task Feasibility for Large Language Models." Under review in ACL Rolling Review. Available in arXiv. [Link] [Code].

[5] Wenbo Zhang, Tong Wu, Yunlong Wang, Yong Cai, and Hengrui Cai. (2023) "Towards Trustworthy Explanation: On Causal Rationalization." International Conference on Machine Learning (ICML), vol. 202, pp. 41715-41736. [Link][Code].

Reinforcement Learning & Bandits

[6] Wenbo Zhang and Hengrui Cai. (2025) "Where to Intervene: Action Selection in Deep Reinforcement Learning." Transactions on Machine Learning Research (TMLR). [Link][Code].

[7] Jiayi Wang*, Wenbo Zhang*, Wenzhuo Zhou, and Hengrui Cai. (2024) "PACE: Pessimistic Adaptive Contextual Bandits for Dynamic Optimal Policy Detection." Submitted to JASA: Special Issue on Statistical Science in Artificial Intelligence, December 2024.

Other Machine Learning Methodology

[8] Eardi Lila, Wenbo Zhang, and Swati Rane. (2024) "Interpretable Discriminant Analysis for Functional Data Supported on Random Nonlinear Domains.", Journal of the Royal Statistical Society Series B (JRSSSB), vol. 86, no. 4, pp. 1013–1044. [Link].

[9] Lars Van Der Laan, Wenbo Zhang, and Peter Gilbert. (2023) "Nonparametric Estimation of the Causal Effect of a Stochastic Threshold-based Intervention." Biometrics, vol. 70, no. 2, pp. 1014-1028. [Link].

Industry Experience

Amazon, Applied Scientist Intern *Summer 2025*
Mentored by Prof. Rui Song, Prof. Sheng Wang and Prof. Hengrui Cai
Topic: Reward modeling for LLM alignment.

- Investigate efficient and effective list-wise feedback learning for reward modeling.
- Design a multi-objective reasoning reward model training method.

Meta, Central Applied Science Team, Research Scientist Intern *Summer 2024*
Mentored by Dr. Wenyu Chen
Topic: Difficulty Quantification for Large Language Models (LLM) Benchmark.

- Developed difficulty metrics (P-correct) as a fine-grained difficulty score for individual prompts.
- Designed LLM-tagging-based difficulty metrics and analyzed a wide range of open-source benchmark datasets.

IQVIA, Advanced Analytics, Machine Learning Research Intern *Summer 2022*
Mentored by Dr. Tong Wu and Dr. Yunlong Wang
Topic: Interpretable Neural Sequence Prediction Models

- Developed a selective rationalization approach for language models (BERT) to explain the predictions by leveraging two causal desiderata, non-spuriousness, and efficiency.
- Applied the method to real-world text and Electronic Health Records (EHR) datasets.

Research Experience

Self-Selective Heterogeneous Reward Model *2025.3-present*
Department of Statistics, University of California, Irvine
Advised by Prof. Hengrui Cai

- Investigated the necessity of rationales in preference judgments and reward modeling for LLM alignment.
- Developed a heterogeneous mixture reward model that selectively leverages self-generated rationales to infer response reward.

RLHF for Large Language Model Alignment *2024.2-2025.5*
Department of Statistics, University of California, Irvine
Advised by Prof. Wenzhuo Zhou and Prof. Hengrui Cai

- Designed controlled experiments to understand the performance gap between offline and iterative direct preference optimization algorithms.

- Designed an offline optimization method that distills knowledge from an explicit preference model to the policy model, achieving alignment without the computational overhead of iterative methods.
- Implemented a max-minimax algorithm to utilize importance ratios and pessimism to mitigate overoptimization in offline alignment.
- Developed calibrated ensemble methods integrating heterogeneous preference models to prevent reward hacking.

Uncertainty Quantification for Large Language Models *2023.9-2024.10*
Department of Statistics, University of California, Irvine
Advised by Prof. Hengrui Cai

- Developed an infeasible benchmark to assess LLMs' refusal capabilities and self-confidence.
- Fine-tuned models to enhance their refusal ability in terms of infeasible tasks.

Reinforcement Learning with High-Dimensional Action Space *2023.1-2024.9*
Department of Statistics, University of California, Irvine
Advised by Prof. Hengrui Cai

- Designed a variable selection method based on conformal inference to find the sufficient and necessary action set from offline trajectory data.
- Developed a hard mask strategy to incorporate offline variable selection results with deep policy/value networks to make online learning more efficient with less spurious features.

Functional Data Analysis for Neuroimaging Diagnosis *2020.9-2023.4*
Department of Biostatistics, University of Washington Seattle
Advised by Prof. Eardi Lila

- Developed a functional penalized regression method over two-dimensional manifolds with a smooth surface penalty; proposed an iterative optimization algorithm to solve this problem

Correlation Study of Antibody Markers with Causal Inference *2019.12-2021.1*
Fred Hutchinson Cancer Research Cente
Advised by Prof. Peter Gilbert

- Helped to develop a non-parametric model to estimate the immune response threshold of risk.
- Used SuperLearner (ensemble models) to predict individual case/control status defined by each endpoint and achieve good classification accuracy.

Profession Activity

Conference Reviewer:
ICLR 2025, AISTATS 2025, ICML 2024; NeurIPS 2024

Journal Reviewer:
Statistical Analysis and Data Mining ($\times$ 1), Journal of Applied Statistics($\times$ 1)

Skills

- **Programming**: Python, PyTorch, R, SQL, Linux, Matlab
- **Tools**: Git, AWS