Wenbo Zhang

PhD Candidate of Statistics

(+1) 9498678694
wenbz13@uci.edu
Website

Research Interest

Uncertainty Quantification, Causality, Large Language Models Alignment, Reinforcement Learning

Education

2021–present **PhD of Statistics**, *University of California, Irvine*
Adviser: Prof. Hengrui Cai

2019–2021 **Master of Science, Biostatistics**, *University of Washington*

2015–2019 **Bachelor of Science, Applied Mathematics**, *Xi'an Jiaotong-Liverpool University*

Fellowships & Awards

2021 **School of Public Health's Outstanding MS Student Award**, awarded to one master student in Department of Biostatistics every year, *University of Washington*.

2020 **UW Summer Institutes Scholarship**, *University of Washington*.

2018 **University Academic Achievement Award**, awarded to 10% of all undergraduates, *XJTLU*.

Publications & Preprints

2024 **On Eliminating Redundant Actions in Reinforcement Learning via Knockoffs**
Wenbo Zhang and Hengrui Cai
In Submission

2024 **Defining Boundaries: A Spectrum of Task Feasibility for Large Language Models**
Wenbo Zhang , Zihang Xu and Hengrui Cai
https://arxiv.org/abs/2408.05873

2024 **Interpretable Discriminant Analysis for Functional Data Supported on Random Nonlinear Domains**
Eardi Lila, Wenbo Zhang , and Swati Rane
Journal of the Royal Statistical Society Series B

2023 **Towards Trustworthy Explanation: On Causal Rationalization**
Wenbo Zhang , Tong Wu, Yunlong Wang, Yong Cai, and Hengrui Cai
International Conference on Machine Learning (ICML), 2023

2022 **Nonparametric Estimation of the Causal Effect of a Stochastic Threshold-based Intervention**
Lars Van Der Laan, Wenbo Zhang , and Peter Gilbert
Biometrics

2021 **Finding Atrophy Patterns of Grey Matter Through Orthonormal Non-negative Factorization**
Wenbo Zhang , Kwun Chuen Gary Chan, Dean Shibata, and David Haynor
SPIE Medical Imaging

2021 **A New Convolutional Neural Network Architecture for Automatic Segmentation of Overlapping Human Chromosomes**
Sifan Song, Tianming Bai, Yanxin Zhao, Wenbo Zhang , Chunxiao Yang, Jia Meng, Fei Ma, and Jionglong Su
Neural Processing Letters

2018 **Chromosome Classification with Convolutional Neural Network Based Deep Learning**
Wenbo Zhang , Sifan Song, Tianming Bai, Yanxin Zhao, Fei Ma, Jionglong Su, and Limin Yu
International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)

Collaboration Papers

2023 **Antibody Correlates of Protection From Severe Respiratory Syncytial Virus Disease in a Vaccine Efficacy Trial**
Youyi Fong, Ying Huang, Bhavesh Borate, Lars Wim and Paul van der Laan, Wenbo Zhang , Lindsay N. Carpp, Iksung Cho, Greg Glenn, Louis Fries, Raphael Gottardo and Peter B. Gilbert
Open Forum Infectious Diseases

2021 **Immune Correlates Analysis of the mRNA-1273 Covid-19 Vaccine Efficacy Clinical Trial**
Peter Gilbert, David montefiori, Adrian Mcdermott, Youyi Fong, David Benkeserw et al.
Science

Industry Experience

June,2024 – Dec,2024 **Research Scientist Intern**
Meta, *Menlo Park, CA*
- Developed difficulty metrics for evaluating Large Language Models and conducted an analysis across a wide range of open-source benchmark datasets at both the prompt and benchmark levels.

June,2022 – Sep,2022 **Machine Learning Research Intern**
IQVIA, *Plymouth Meeting, PA (Remote)*
- Developed a novel selective rationalization approach based on large language models to explain the predictions by leveraging two causal desiderata, non-spuriousness, and efficiency for Natural Language Processing (NLP) and Electronic Health Records (EHR) datasets

Research Experience

April,2024 – present **Reinforcement Learning From Human Feedback for Large Language Model Alignment**
Department of Statistics, University of California Irvine, Irvine, CA
- Understand the performance gap between offline and iterative direct preference optimization algorithms
- Designed an efficient iterative algorithm incorporating reward margin reweighting.

Sep,2023 – present **Uncertainty Quantification with Large Language Model Generations**
Department of Statistics, University of California Irvine, Irvine, CA
- Developed an infeasible benchmark to assess LLMs' refusal capabilities and confidence elicitation.
- Fine-tuned models to enhance their refusal ability.

Jan,2023 – present **Reinforcement Learning with High-Dimensional Action Space**
Department of Statistics, University of California Irvine, Irvine, CA
- Unitized variable selection method to find the sufficient and necessary action set from offline data and make online learning more efficient with less spurious features

Sep,2020 – Mar,2021 **Functional Data Analysis for Neuroimaging Diagnosis**
Department of Biostatistics, University of Washington, Seattle, WA
- Developed a functional penalized regression method over two-dimensional manifolds with a smooth surface penalty; proposed an iterative optimization algorithm to solve this problem

Jun,2020 – Sep,2020 **Correlation Study of Antibody Markers with Causal Inference**
Fred Hutchinson Cancer Research Center, Seattle, WA
- Helped to develop a non-parametric model based on Causal Inference techniques to estimate immune response threshold of risk

Professional Activity

Conference Reviewer: ICML 2024; Neurips 2024

Skills

Programming Languages	Python, PyTorch, R, SQL, Linux, Matlab