# Zihao Wang

920 E 58th St, Office 408, Chicago, IL 60637 • wangzh@uchicago.edu • (312) 394-0229

## Skills & Interests

- Interpretability and robustness for foundation models, AI safety, AI Alignment
- Causal discovery & inference, causality for ML, invariant representation learning
- Bayesian modeling for exploratory data analysis, prediction and inference
- Model diagnostics, hypothesis testing, experiment design
- Programming skills: Python(advanced), R(advanced), Matlab(advanced), C(intermediate), PyTorch (advanced), TensorFlow (intermediate)

## EDUCATION

**University of Chicago** | PhD program in Statistics | GPA: 3.86/4.00 Sep 2020 - present
**University of Chicago** | B.S., Computational and Applied Mathematics | GPA 3.95/4.00 Sep 2019
**Selected Courses** (all A): Linear Models, GLM, Nonparametric Inference, Math Statistics | Convex Optimization, Deep Learning, Advanced NLP, Stochastic Simulation | Theory of Algorithm, Measure Theoretic Probability

## Publications

- **Zihao Wang**, Victor Veitch (2022). A Unified Causal View of Domain Invariant Representation Learning (arXiv:2208.06987). Presentation at ICML SCIS workshop 2022, submitted to ICLR 2023
- Peter Carbonetto, Abhishek Sarkar, **Zihao Wang**, Matthew Stephens (2021), Non-negative matrix factorization algorithms greatly improve topic model fits (arXiv:2105.13440).

## RESEARCH EXPERIENCE

**Concept Control for Large Text-to-Image Diffusion Model** Aug 2022 – present
Advisor: **Prof. Victor Veitch**, Department of Statistics University of Chicago, Google Brain

- Goal is to study and solve the problem of concept entanglement in text-to-image diffusion models
- Found instances where text-guided models fail to activate concepts in isolation and show biases
- Discovered heuristic ways for concept control and working to develop principled methods

**A Unified Causal View of Domain Invariant Representation Learning** Jan 2022-Sep 2022
Advisor: **Prof. Victor Veitch**, Department of Statistics University of Chicago, Google Brain

- Studied the problem of Domain Shifts in Machine Learning through a novel Causal Framework
- Characterized the relationships among various domain-invariant representation learning methods: data augmentation, distributional-invariance learning and invariant risk minimization, and give recommendations for which methods to use in practice
- Performed experiments on synthetic and large-scale problems with domain shifts with Pytorch
- Poster presentation at ICML SCIS workshop 2022 & full paper submission to ICLR 2023

**Fast Algorithm for Nonnegative Matrix Factorization (NMF)** June 2021 – present
Advisor: **Prof. Matthew Stephens**, Department of Statistics and Human Genetics, University of Chicago

- Goal is to develop a very fast algorithm for nonnegative matrix factorization by adding extra reasonable data assumptions. We start from "anchor word"-based algorithm, which is fast and provably correct, but is not robust in the presence of noise.
- Identified why "anchor word"-based algorithm fails, and designed a few heuristic fixes that improve practical performance.
- Currently working on more principled statistical approaches in place of the heuristic fixes

**Nonnegative Matrix Factorization (NMF) on count data** June 2018 – August 2020
Advisor: **Prof. Matthew Stephens**, Department of Statistics and Human Genetics, University of Chicago

- Proposed Empirical Bayes approach to nonnegative matrix factorization for count data
- Implemented methods in in R package ebpmf, and apply to large-scale genetics and text datasets, showing improvement in interpretability

## Course Projects

**Statistics Clinics** Sep 2021 - Present

- Led a statistical consulting team at UChicago, helping clients from biology to social sciences.
- Helped clients turn scientific questions into statistical hypotheses; design tests for structured data.

**Advanced Natural Language Processing** (repo: nlp_projects) March 2019 – June 2019
Instructor: **Prof. Kevin Gimpel,** Toyota Technological Institute at Chicago

- Implemented Sentiment Analysis with Word-Embedding and Attention function in PyTorch.
- Implemented Hidden Markov Model for Structured Prediction through Viterbi algorithm, Greedy algorithm, Beam Search algorithm, and Gibbs Sampling.
- Built an unsupervised segmenter for English text with nonparametric Bayes using Gibbs Sampling.