# Wenhao PAN

wenhaop.github.io | linkedin.com/in/wenhao-pan-uw | pan20001102@gmail.com | (510) 646-3809

## EDUCATION

**University of Washington, Seattle** — Seattle, WA
Ph.D. in **Statistics** (GPA: 3.94/4.00) — 09/2023 - Present
- Highlighted courses: Statistical Inference I & II, Discrete & Continuous Stochastic Processes, Multiple Testing and Modern Inference, High-dimensional Statistical Learning, Machine Learning for Big Data, Bandits, Measure Theory, Stochastic Calculus (In Progress).

**University of California, Berkeley** — Berkeley, CA
B.A. in **Statistics** and B.A. in **Computer Science** (*Summa cum laude*) — 08/2019 - 05/2023

## RESEARCH EXPERIENCE

**University of Washington, Seattle** | **Advisor: Tyler McCormick, Zaid Harchaoui** — Seattle, WA
**Project:** Incorporating Network Interference into Performative Prediction — 08/2024 - Present
- Replaced the Stable Unit Treatment Values Assumption in a causal experiment under performative prediction with network interference.

**University of Washington, Seattle** | **Advisor: Daniela Witten** — Seattle, WA
**Project:** Improving the Data Thinning Algorithm through the Rank-transformed subsampling Algorithm — 08/2023 - 02/2024
- Improve the statistical power of Data Thinning algorithm while maintaining type I error control through Rank-transformed subsampling.

**Berkeley Artificial Intelligence Research Lab** | **Advisor: Anil Aswani** — Berkeley, CA
**Project:** Improving the Numerical Computation of a Novel Nonnegative Tensor Completion Approach — 05/2022 - Present
- Accelerated a nonnegative tensor completion approach proposed in the NeurIPS 2022 paper "Nonnegative Tensor Completion via Integer Optimization" to increase its applicability to healthcare, computer vision, and other domains.
- Developed 10 variants by optimizing data structures, accelerating gradient descent, and applying blended pairwise conditional gradients.
- Reduced the computation time of using aforementioned approach to demosaic a $90 \times 60$ pixel image from 11,300 seconds to 1,500 seconds.

**Lawrence Berkeley National Laboratory** | **Advisor: Haichen Wang** — Berkeley, CA
**Project:** Developing a Generative Neural Network for Probability Distribution Modeling in Particle Physics — 10/2022 - Present
- Improved a conditional normalizing flow model to simulate the correlated and asymmetric distribution of particle detector responses.

**Project:** Developing a Graph Transformer Neural Network for Regression in Particle Physics — 01/2022 - 10/2022
- Improved the deep-learning model proposed in the paper "A Holistic Approach to Predicting Top Quark Kinematic Properties with the Covariant Particle Transformer" for predicting the kinematic quantities of Higgs bosons in the $t\bar{t}H$ production samples.
- Fit a PyTorch Graph Transformer model with 13 million parameters onto 7 million samples using the supercomputer Cori.
- Addressed the model's underprediction of the transverse momentum (pT) of Higgs bosons by reweighting its loss function.
- Increased the percentage of Higgs bosons with both true and predicted pT in a high pT interval from 26% to 44%.

**Oski Lab** | **Advisor: Cyrus Dioun** — Berkeley, CA
**Project:** Using Natural Language Processing and Deep Learning for Product Classification — 02/2021 - 10/2022
- Hand-coded 3,100 cannabis products to create a labeled dataset to build multi-label classification deep learning models.
- Optimized a Keras TextCNN model through hyperparameter tuning to achieve a 93.7% average F1 score for five labels in the testing set.
- Fine-tuned a PyTorch BERT model with Hugging Face to achieve a 95.3% average F1 score for five labels in the testing set.
- Recognized as the most significant student contributor to this project in the acknowledgment of its working paper.

**Project:** Racial Stratification in the Washington State Cannabis Industry — 05/2021 - 06/2021
- Compared ownership demographic data with revenue and enforcement data to examine relative diversity in the cannabis industry.
- Validated the project in Python with Pandas, NumPy, and Matplotlib, and added extra functionality through GitHub.

## PUBLICATIONS

- **Pan, W.**, Aswani, A. and Chen, C., Accelerated Nonnegative Tensor Completion via Integer Programming. Frontiers in Applied Mathematics and Statistics, 9, p.1153184.

## TEACHING EXPERIENCE

**University of Washington, Seattle** | **Teaching Assistant** — Seattle, WA
- STAT 516, Stochastic Modeling of Scientific Data — 09/2024 - Present
- STAT 390, Statistical Methods in Engineering and Science — 06/2024 - 08/2024
- CSE 416, Introduction to Machine Learning — 03/2024 - 06/2024
- STAT 180, Introduction to Data Science — 01/2024 - 03/2024

**University of California, Berkeley** | **Academic Student Employee** — Berkeley, CA
- STAT 151A, Linear Modeling: Theory and Applications, Grader — 01/2023 - 05/2023
- STAT 151A, Linear Modeling: Theory and Applications, Grader — 08/2022 - 12/2022
- DATA C140, Probability for Data Science, Group Tutor — 01/2022 - 05/2022

## EXTRACURRICULAR ACTIVITIES

**Student Association of Applied Statistics** — Berkeley, CA
**Department:** Career Exploration, Data Consulting, Education — 09/2020 - 05/2023
- Conducted regression analysis on the relationship between podcasters' language styles and popularity in a project sponsored by Tencent.
- Mentored potential statistics or data science major students and recited two-hour lectures on topics in data science and applied statistics.

## SKILLS

- Programming: Python, R, Java, C++, MATLAB, SQL, ROOT
- Packages & Tools: NumPy, Pandas, Matplotlib, Scikit-Learn, PyTorch, PySpark, HuggingFace, Latex, Git, JupyterLab, Anaconda