ZHANYU WANG

Address: 2400 W El Camino Apt 502, Mountain View, CA, 94040   **Phone**: (765) 409-6928
Email: wzy1993xt@gmail.com   **Website**: https://www.stat.purdue.edu/~wang4094/

EDUCATION

Purdue University, West Lafayette, IN, U.S.   08/2018 - 12/2023
- Ph.D. in Statistics, Department of Statistics. Advisors: Dr. Jordan Awan and Dr. Guang Cheng.
- Research Area: Differential Privacy. Optimization in Deep Learning. Meta-learning.

Chinese Academy of Sciences, Beijing, China   09/2015 - 07/2018
- M.S. in Statistics, Academy of Mathematics and Systems Science. Advisor: Dr. Lei Li.

Peking University, Beijing, China   09/2011 - 07/2015
- B.S. in Statistics, School of Mathematics Sciences.

PUBLISHED PAPERS

- Awan, J., **Wang, Z.** "Simulation-based Confidence Intervals and Hypothesis Tests for Privatized Data." *Journal of the American Statistical Association. 2025.*

- Wang, C., **Wang, Z.**, Sun, W., Cheng, G. "Online Regularization for High-Dimensional Dynamic Pricing Algorithms." *Journal of the American Statistical Association. 2023.*

- **Wang, Z.**, Zhang, X., Yun, H., Teo, C., Chilimbi, T. "MICO: Selective Search with Mutual Information Co-training." *COLING, 2022* (Oral presentation).

- **Wang, Z.**, Honorio, J. "The Sample Complexity of Meta Sparse Regression." *AISTATS, 2021*.

- Chao, S., **Wang, Z.**, Xing, Y., Cheng, G. "Directional Pruning of Deep Neural Networks." *NeurIPS, 2020*.

- Li, W., **Wang, Z.**, Zhang, Y., Cheng, G. "Variance Reduction on Adaptive Stochastic Mirror Descent." *Machine Learning, 2022*. (Conference version in *NeurIPS workshop, OPT 2020*.)

- Wang, A., **Wang, Z.**, Li, Z., Li, L. "BAUM: Improving genome assembly by adaptive unique mapping and local overlap-layout-consensus approach." *Bioinformatics, 34.12 (2018), 2019-2028*.

SUBMITTED PAPERS

- **Wang, Z.**, Cheng, G., Awan, J. "Differentially Private Bootstrap: New Privacy Analysis and Inference Strategies." *Journal of Machine Learning Research.* (Reject and resubmit.)

TALKS AND POSTERS

Theory and Practice of Differential Privacy (TPDP 2023), Boston, U.S.   09/2023
- (Poster) De-Biased Parametric Bootstrap Inference on Privatized Data

Joint Statistical Meetings (JSM 2023), Toronto, Canada   08/2023
- (Talk) De-Biased Parametric Bootstrap Inference on Privatized Data

Statistics and Optimization in Data Science Workshop, Purdue University   05/2023
- (Poster) Differentially Private Bootstrap: New Privacy Analysis and Inference Strategies

Midwest Machine Learning Symposium (MMLS 2023), Chicago, U.S.   05/2023
- (Poster) Differentially Private Bootstrap: New Privacy Analysis and Inference Strategies

Eighth Bayesian, Fiducial, and Frequentist conference (BFF8), Cincinnati, U.S.   05/2023
- (Poster) Simulation-based Confidence Intervals and Hypothesis Tests for Privatized Data

Future Leaders Summit, Ann Arbor, Michigan, U.S.   04/2023
- (Talk) Differentially Private Bootstrap: New Privacy Analysis and Inference Strategies

Machine Learning for Business (MGMT 47500), Purdue University   02/2023
- (Talk) MICO: Selective Search with Mutual Information Co-training

Joint Statistical Meetings (JSM 2022), Washington D.C., U.S.   08/2022
- (Poster) Differentially Private Bootstrap: New Privacy Analysis and Inference Strategies

ACADEMIC SERVICES

Action editor
- Journals: Transactions on Machine Learning Research (2024 -).

Reviewer
- Conferences: ICLR (2023, 2024), ICML (2022, 2023, 2024), NeurIPS (2021, 2022, 2023)

- Journals: Journal of the American Statistical Association. Scientific reports. Statistical analysis and data mining. Journal of Data Science.

Honors

• Honorable mention of the Student Paper Competition in Joint Statistical Meetings (JSM)	2024
• Top (10%) reviewers of NeurIPS	2023
• Bilsland Dissertation Fellowship. Purdue University	2023
• Second Place Winner Poster Award. Statistics and Optimization in Data Science Workshop.	2023
• Outstanding Poster Award (5/141). Midwest Machine Learning Symposium (MMLS).	2023
• Travel Grant. Purdue Graduate Student Government.	2023
• Top (10%) reviewers of NeurIPS	2022
• Outstanding (10%) reviewers of ICML	2022
• Ross Fellowship. Purdue University	2018
• China National Scholarship (Graduate)	2017
• Silver Medal, National Final, Chinese Mathematics Olympiad. Chinese national training team for IMO.	2011

Teaching and Consulting

Purdue University, *Research Assistant*, West Lafayette, US — 08/2022 - 08/2023
- Work with Dr. Jordan Awan on differential privacy.

Purdue University, *Teaching Assistant*, West Lafayette, US — 08/2021 - 05/2022
- STAT 303 Coordinator and Head TA. Teaching probability and statistics to undergraduate students.

Purdue University, *Data Science Consultant*, West Lafayette, US — 01/2021 - 05/2021
- We propose a method for the defense of various adversarial attacks with meta-learning.

Purdue University, *Teaching Assistant*, West Lafayette, US — 08/2019 - 12/2020
- STAT 225 Instructor. Teaching probability to undergraduate students.

Professional careers

Meta, *Research Scientist*, Menlo Park, US — 12/2023 - Now

Meta, *Software Engineer Intern (Machine Learning Engineer track)*, Menlo Park, US — 05/2022 - 08/2022
- We build a learning-based system for Ads Pacing using primal-dual algorithms. Our next-generation pacer model can adapt to market change promptly and also has strong explainability in terms of marginal cost.
- We implement the pacer using C++ and test it in the production environment of Meta; By experiments on 5% real-world traffic of the company, we have observed a significant increase (0.4%) in Ads value.

Amazon, *Applied Scientist Intern*, Seattle, US — 05/2021 - 08/2021
- We propose a selective search framework based on Mutual Information Co-training (MICO) (clustering documents to groups by their similarity, and searching each query only within its most relevant groups). We reduce the search cost to 5% with achieving 99% accuracy compared to searching on all documents.
- MICO is an end-to-end learning model. Its objective function is the mutual information between the two group indices of a query and its related document, both of which are outputs of trainable neural networks.
- In my implementation, the model takes BERT representation of a sentence (query or document title) as input, and it can be efficiently trained on huge dataset (hundreds of GB) with BERT also being finetuned.
- The paper of MICO is accepted by COLING 2022 as Oral Presentation (10%).

Skills

- **Programming Languages**: experienced in Python, R, and C++, familiar with C, Perl, Shell, SQL, etc.
- **Softwares & Systems**: PyTorch, TensorFlow (GPU & TPU), Git, Docker, Linux, etc.