YAN SUN

1815 John F Kennedy Blvd, Philadelphia, PA 19103
765-775-3724, sunyan1@wharton.upenn.edu

Website | Google Scholar | Github | LinkedIn

EDUCATION

Purdue University *August 2017 - May 2022*
Ph.D. in Statistics *West Lafayette, IN*

Dissertation: Sparse Deep Learning and Stochastic Neural Network

Advisor: Dr. Faming Liang

M.Sc. in Joint Statistics and Computer Science

Zhejiang University *September 2013 - June 2017*
B.S. in Mathematics and Applied Mathematics *Hangzhou, China*

Quishi Pursuit Science Class of Chu Kochen Honors College

EMPLOYMENT

University of Pennsylvania *May 2023 - Now*
Postdoctoral Researcher *Philadelphia, PA*

Research on machine learning model calibration and uncertainty quantification

Amazon *June 2022 - March 2023*
Applied Scientist I *Seattle, WA*

Apply multi-task neural network model to Import Fee Deposit (IFD) prediction

Refine model validation logic to automatically verify model performance

TEACHING EXPERIENCE

Purdue University

Instructor: STAT 225, Introduction to Probability Models *August 2018 - May 2019*

Teaching Assistant: STAT 301, Elementary Statistical Methods *August 2017 - May 2018*

RESEARCH INTERESTS

Uncertainty Quantification for Machine Learning Models: calibration, asymptotic theory.

Sparse Deep Learning: consistency, high dimensional variable selection, network pruning.

Bayesian Statistics: Bayesian neural network, posterior concentration, MCMC.

PREPRINTS OR SUBMITTED MANUSCRIPTS

1. Gao, Zijun, **Yan Sun**. Trustworthy Evaluation of Generative AI Models. Paper submitted to *International Conference on Machine Learning*.
2. Chao, Patrick, **Yan Sun**, Edgar Dobriban, Hamed Hassani. Watermarking Language Models with Error Correcting Codes. Paper submitted to *International Conference on Machine Learning*.

3. **Sun, Yan**, Pratik Chaudhari, Ian Barnett, and Edgar Dobriban. A Confidence Interval for the ℓ_2-Expected Calibration Error. arXiv:2408.08998 (2024). Paper submitted to *Journal of the American Statistical Association.*

4. **Sun, Yan**, Faming Liang. Uncertainty Quantification for Large-Scale Deep Neural Networks via Post-StoNet Modeling. Paper submitted to *Statistica Sinica.*

PUBLICATIONS

1. Zhang, Mingxuan, **Yan Sun**, and Faming Liang. Magnitude Pruning of Large Pretrained Transformer Models with a Mixture Gaussian Prior. *Journal of Data Science* (2024) in press.

2. Liang, Faming, Sehwan Kim, and **Yan Sun**. Extended fiducial inference: toward an automated process of statistical inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology* (2024): qkae082.

3. Dong, Tianning, **Yan Sun**, and Faming Liang. Deep network embedding with dimension selection. *Neural Networks* 179 (2024): 106512.

4. Zhang, Mingxuan, **Yan Sun**, and Faming Liang. Sparse deep learning for time series data: theory and applications. *Advances in Neural Information Processing Systems* 36 (2024).

5. Liang, Siqi, **Yan Sun**, and Faming Liang. Nonlinear sufficient dimension reduction with a stochastic neural network. *Advances in Neural Information Processing Systems* 35 (2022): 27360-27373.

6. **Sun, Yan**, and Faming Liang. A kernel-expanded stochastic neural network. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.2 (2022): 547-578.

7. **Sun, Yan**, Qifan Song, and Faming Liang. Learning sparse deep neural networks with a spike-and-slab prior. *Statistics & probability letters* 180 (2022): 109246.

8. **Sun, Yan**[1], Qifan Song[1], and Faming Liang. Consistent sparse deep learning: Theory and computation. *Journal of the American Statistical Association* 117.540 (2022): 1981-1995.

9. **Sun, Yan**, Wenjun Xiong, and Faming Liang. Sparse deep learning: A new framework immune to local traps and miscalibration. *Advances in Neural Information Processing Systems* 34 (2021): 22301-22312.

10. Song, Qifan, **Yan Sun**, Mao Ye, and Faming Liang. Extended stochastic gradient Markov chain Monte Carlo for large-scale Bayesian variable selection. *Biometrika* 107.4 (2020): 997-1004.

11. Ye, Mao[1], and **Yan Sun**[1]. Variable selection via penalized neural network: a drop-out-one loss approach. *International Conference on Machine Learning*. PMLR, 2018.

HONORS AND AWARDS

IDEAS AI X Science Postdoc Program Award, University of Pennsylvania. *November 2024*

William J. Studden Publication Award, Purdue Department of Statistics. *April 2022*

Bilsland Dissertation Fellowship, Purdue Depeartment of Statistics *April 2021*

Virgil Anderson and Gloria Fischer Graduate Fellowship, Purdue Department of Statistics *April 2021*

William J. Studden Publication Award, Purdue Department of Statistics. *April 2021*

ICML Travel Award *May 2018*

RESEARCH TALKS AND POSTERS

WNAR/IBS/Graybill 2024 Conference, Fort Collins, Colorado, June 2024

NYU Department of Technology, Operations, and Statistics Seminar, New York, January 2024

NeurIPS 2022 poster presentation, New Orleans, Louisiana, December 2022

NeurIPS 2021 poster presentation, Virtual, December 2021

GSO Student Seminar Series, Purdue, March 2021

ICML 2018, Stockholm, Sweden, July 2018

PROFESSIONAL SERVICES

Conference organization

Session chair and organizer. Invited session on "Uncertainty Quantification for Large Language Models (LLMs)" at JSM 2025. Invited speakers: John Cherian, Ying Jin, Jiaxin Zhang

Conference Reviewer

ICML, NeurIPS, ICLR, AAAI.

Journal Reviewer

Journal of the American Statistical Association (JASA), Electronic Journal of Statistics (EJS), IEEE Transactions on Neural Networks and Learning Systems (TNNLS), Scientific Reports.