GANGHUA WANG

Email: ganghua@uchicago.edu

Address: Data Science Institute, 5460 S University Ave, Room 202, Chicago, IL, 60615

Research Interests: Trustworthy AI, Machine learning theory and applications, Large generative models

PROFESSIONAL EXPERIENCE

University of Chicago, Data Science Institute *Sept. 2024 - present*
Postdoctoral Researcher *Chicago, IL*
Co-hosted by Prof. Haifeng Xu and Prof. Bo Li

EDUCATION

University of Minnesota, Twin Cities *Aug. 2019 - Aug. 2024*
Ph.D. in Statistics *Minneapolis, MN*
Advised by Prof. Jie Ding and co-advised by Prof. Yuhong Yang

Peking University *Sept. 2015 - July 2019*
B.S. in Statistics, Minor in Economics *Beijing, China*

PUBLICATIONS

Published

* indicates equal contributions

[1] An Luo, Xun Xian, Jin Du, Fangqiao Tian, **Ganghua Wang**, Ming Zhong, Shengchun Zhao, Xuan Bi, Zirui Liu, Jiawei Zhou, et al. "AssistedDS: Benchmarking How External Domain Knowledge Assists LLMs in Automated Data Science". In: *Proc. EMNLP* (2025).

[2] Xun Xian, **Ganghua Wang**, Xuan Bi, Rui Zhang, Jayanth Srinivasa, Ashish Kundu, Charles Fleming, Mingyi Hong, and Jie Ding. "On the Vulnerability of Applying Retrieval-Augmented Generation within Knowledge-Intensive Application Domains". In: *Proc. ICML* (2025).

[3] **Ganghua Wang**, Ali Payani, Myungjin Lee, and Ramana Kompella. "Mitigating Group Bias in Federated Learning: Beyond Local Fairness". In: *Trans. Mach. Learn. Res.* (2024). [pdf].

[4] **Ganghua Wang***, Xun Xian*, Jayanth Srinivasa, Ashish Kundu, Xuan Bi, Mingyi Hong, and Jie Ding. "Demystifying Poisoning Backdoor Attacks from a Statistical Perspective". In: *Proc. ICLR* (2024). [pdf].

[5] Xun Xian, **Ganghua Wang**, Xuan Bi, Jayanth Srinivasa, Ashish Kundu, Mingyi Hong, and Jie Ding. "RAW: A Robust and Agile Plug-and-Play Watermark Framework with Provable Guarantees". In: *Proc. NeurIPS* (2024). [pdf].

[6] Enmao Diao*, **Ganghua Wang***, Jie Ding, Yuhong Yang, and Vahid Tarokh. "Pruning deep neural networks from a sparsity perspective". In: *Proc. ICLR* (2023). [pdf].

[7] Gen Li, **Ganghua Wang**, and Jie Ding. "Provable Identifiability of Two-Layer ReLU Neural Networks via LASSO Regularization". In: *IEEE Trans. Inf. Theory* 69.9 (2023), pp. 5921–5935. DOI: 10.1109/TIT.2023.3274152.

[8] **Ganghua Wang**, Jie Ding, and Yuhong Yang. "Regression with Set-Valued Categorical Predictors". In: *Statistica Sinica* 33.4 (2023), pp. 2545–2560. DOI: 10.5705/ss.202021.0332.

[9] Xun Xian, **Ganghua Wang**, Jayanth Srinivasa, Ashish Kundu, Xuan Bi, Mingyi Hong, and Jie Ding. "A Unified Framework for Inference-Stage Backdoor Defenses". In: *Proc. NeurIPS* (2023). [pdf].

[10] Xun Xian*, **Ganghua Wang***, Jayanth Srinivasa, Ashish Kundu, Xuan Bi, Mingyi Hong, and Jie Ding. "Understanding backdoor attacks through the adaptability hypothesis". In: *Proc. ICML* (2023). [pdf].

Under review

[11] **Ganghua Wang**, Yuwei Chen, and Haifeng Xu. "Exponential Convergence of Probabilistic Bisection Algorithm with Noisy Labels". In: *Proc. ICLR* (2025).

[12] **Ganghua Wang**, Zhaorun Chen, Bo Li, and Haifeng Xu. "Cer-Eval: Certifiable and Cost-Efficient Evaluation Framework for LLMs". In: *Proc. ICLR* (2025).

[13] **Ganghua Wang** and Jie Ding. "Subset Privacy: Draw from an Obfuscated Urn". In: *arXiv preprint* (2025). [pdf].

[14] **Ganghua Wang**, Oliver R. Wang, Bo Li, and Haifeng Xu. "Extrapolating Large Models from the Small: Optimal Learning of Scaling Laws". In: *Proc. ICLR* (2025).

[15] **Ganghua Wang**, Yuhong Yang, and Jie Ding. "Model Privacy: A Framework to Understand Model Stealing Attack and Defense". In: *Submitted to JRSSB, under major revision* (2025).

[16] Yangjianchen Xu, Qinglong Tian, and **Ganghua Wang**. "Semiparametric Regression Analysis of Right-Censored Events with Privacy-Preserving Data". In: *J. Am. Stat. Assoc.* (2025).

[17] Wenjing Yang*, **Ganghua Wang***, Jie Ding, and Yuhong Yang. "A Theoretical Understanding of Neural Network Compression from Sparse Linear Approximation". In: *IEEE Trans. Signal Process.* (2025). [pdf].

Manuscript

[18] **Ganghua Wang**, Zhiyuan Tang, and Jie Ding. "Classification with set-valued labels". In: *Preparation* (2025).

[19] **Ganghua Wang**, Yuhong Yang, and Jie Ding. "Regression with Adversarially Perturbed Responses". In: *Preparation* (2025).

HONORS AND AWARDS

Awarded by University of Chicago, Data Science Institute

Faraco Postdoctoral Fellowship for Outstanding Research — *2025*

Mentor of the Year Award — *2025*

Awarded by University of Minnesota, Twin Cities

Doctoral Dissertation Fellowship Finalist — *2023*

School of Statistics Travel Award — *2023*

Summer Research Fellowship — *2020*

School of Statistics First Year Scholarship — *2019*

Awarded by Cisco Systems, Inc.

Cisco Research Graduate Award — *2022*

Cisco Research Fellow — *2022, 2023*

Awarded by Peking University

The Academic Excellence Scholarship (3 times) — *2016 - 2018*

Fang Zheng Scholarship — *2017*

Wu Si Scholarship — *2016, 2018*

Freshman Scholarship — *2015*

TEACHING AND MENTORING

Mentor of Data Science Clinic, University of Chicago

Mapping Human Rights Violations in the Palm Oil Industry — *Spring 2025*

Inclusive Development International
Climate change-induced multi-hazard risks: Landslides *Winter 2024*
University of Rwanda
Characterizing African American Young Adult Novels' Narrative *Autumn 2024*
University of Northern Iowa

Mentor of
Undergraduate Reading Group, School of Statistics, University of Minnesota *2024*

AEOP High School Apprenticeship K-12 Outreach Program to enhance diversity, equity, and inclusion (sponsored by Army Research Office) *2022*

Teaching Assistant, University of Minnesota, Twin Cities
STAT 4102 Theory of Statistics II *Fall 2020*
STAT 3021 Introduction to Probability and Statistics *Spring 2020*
STAT 3011 Introduction to Statistical Analysis *Fall 2019*

SERVICES

Member of
Graduate Student Liaison Committee, *June 2023 - Aug. 2024*
School of Statistics, University of Minnesota

Reviewer of
Journal of the American Statistical Association, IEEE Transactions on Information Theory, Transactions on Machine Learning Research, SIAM Journal on Mathematics of Data Science, AISTATS, AAAI, ICML, ICLR, ICASSP, NeuRIPS, Neural Processing Letter, Machine Learning

SOFTWARE

Python Packages

1. `SubsetPrivacy`: Implement the subset privacy mechanisms proposed in [13] and statistical inference methods for set-valued observations [8].
2. `ModelPrivacy`: Implement the common and proposed model stealing defense/attack strategies on benchmark datasets [15].
3. `FedGFT`: Implement a bias mitigation federated learning algorithm proposed in [3]. It has been incorporated into Flame, an open-source project for federated learning systems, which is developed by Cisco Systems, Inc.
4. `CBD`: Conformal backdoor defense that detects the backdoor inputs in the inference stage with provable guarantees on false positive rate [9].
5. `RAW`: A robust and agile watermarking technique for AI-generated images [5].

PATENTS

1. G. Wang, M. Lee, A. Payani, R. Kompella, "Group Bias Mitigation in Federated Learning Systems," 07/28/2023, US patent #18/227,535

TECHNICAL STRENGTHS

Computer Languages Pyth on, MATLAB, R, SQL

PRESENTATIONS AND TALKS

Guest Lecturer, Clemson University, Clemson, SC	Nov. 2025
Data Science Institute, University of Chicago, Chicago, IL	Nov. 2025
Data Science Institute Research Day, University of Chicago, Chicago, IL	May 2025
Clemson University, Clemson, SC	Oct. 2024
International Conference on Econometrics and Statistics, Beijing, China	July 2024
International Conference on Learning Representations, Vienna, Austria	May 2024
Department of Statistics, Chinese University of Hong Kong, Hong Kong	Apr. 2024
Conference on Neural Information Processing Systems, New Orleans, LA	Dec. 2023
School of Statistics, University of Minnesota, Minneapolis, MN	Nov. 2023
Joint Statistical Meeting, Toronto, Canada	Aug. 2023
International Conference on Econometrics and Statistics, Tokyo, Japan	Aug. 2023
International Conference on Machine Learning, Honolulu, HI	July 2023
Joint Conference on Statistics and Data Science, Beijing, China	July 2023
Center of Statistical Research, Southwestern University of Finance and Economics, Chengdu, China	June 2023
Center for Statistical Science, Peking University, Beijing, China	June 2023
International Conference on Learning Representations, Kigali, Rwanda	May 2023
School of Statistics Student Seminar, Minneapolis, MN	Mar. 2023