Jiachen (Tianhao) Wang

E-Quad, Princeton, NJ – 08540 – USA

website tianhaowang@princeton.edu github 1-6176469585

RESEARCH INTERESTS

I develop principled **data-centric** methodologies to build **trustworthy AI at scale**. Poor data decisions cascade through the entire AI lifecycle, resulting in opaque models with unreliable predictions, massive compute waste on low-quality data, and private information leakage. To this end, I design algorithms for **data attribution** (quantifying data influence), **data curation** (optimizing dataset quality), and **data privacy** (provably preventing data leakage) that are both *theoretically-grounded* and *scalable to foundation models*. My work have been awarded a **best paper honorable mention** and **9 oral/spotlight presentations** at top-tier AI conferences, and has achieved real-world impact through adoption by Google's data team for frontier AI development and an EU-funded data marketplace platform.

EDUCATION

Princeton University — **Princeton, NJ**
PhD Candidate in Electrical and Computer Engineering — *Sept 2021 - Present*
Advisor: **Prof. Prateek Mittal**

Harvard University — **Cambridge, MA**
MEng in Computational Science and Engineering — *Sept 2019 - May 2021*
Thesis: Concurrent Composition Theorem of Differential Privacy
Advisor: **Prof. Salil Vadhan**

University of Waterloo — **Waterloo, ON**
BMath in Computer Science and Statistics — *Sept 2016 - May 2019*

PUBLICATIONS

Note: *NeurIPS, ICLR, and ICML are widely recognized as top-tier conferences in general machine learning. AISTATS is a top-tier conference at the intersection of artificial intelligence and statistics, with a strong emphasis on statistical approaches in machine learning.*

*I have **led or contributed significantly** to the following research papers.*

Capturing the Temporal Dependence of Training Data Influence
Jiachen T. Wang, Dawn Song, James Zou, Prateek Mittal, Ruoxi Jia
ICLR 2025 **(Oral Presentation, top 1.5% among submissions)** {PDF}

Data Shapley in One Training Run
Jiachen T. Wang, Prateek Mittal, Dawn Song, Ruoxi Jia
ICLR 2025 **(Oral Presentation, top 1.5% among submissions)** {PDF}
Outstanding Paper Honorable Mention (6 out of 11,000+ submissions)
Integrated into Google's data efforts for frontier AI development.

GREATS: Online Selection of High-Quality Data for LLM Training in Every Iteration
Jiachen T. Wang, Tong Wu, Dawn Song, Prateek Mittal, Ruoxi Jia
NeurIPS 2024 **(Spotlight Presentation, top 3% among submissions)** {PDF}

Rethinking Data Shapley for Data Selection Tasks: Misleads and Merits
Jiachen T. Wang, Tianji Yang, James Zou, Yongchan Kwon, Ruoxi Jia
ICML 2024 **(Oral Presentation, top 1.5% among submissions)** {PDF}

Efficient Data Valuation for Weighted Nearest Neighbor Algorithms
Jiachen T. Wang, Prateek Mittal, Ruoxi Jia
AISTATS 2024 **(Oral Presentation, top 1.6% among submissions)** {PDF}

DP-OPT: Make Large Language Model Your Privacy-Preserving Prompt Engineer
Junyuan Hong, **Jiachen T. Wang**, Chenhui Zhang, Zhangheng Li, Bo Li, Zhangyang Wang
ICLR 2024 **(Spotlight Presentation, top 3% among submissions)** {PDF}

Privacy-Preserving In-Context Learning for Large Language Models
Tong Wu*, Ashwinee Panda*, **Jiachen T. Wang***, Prateek Mittal
ICLR 2024 {PDF}

Threshold KNN-Shapley: A Linear-Time and Privacy-Friendly Approach to Data Valuation
Jiachen T. Wang, Yuqing Zhu, Yu-Xiang Wang, Ruoxi Jia, Prateek Mittal
NeurIPS 2023 **(Spotlight Presentation, top 3% among submissions)** {PDF}

A Randomized Approach for Tight Privacy Accounting
Jiachen T. Wang, Saeed Mahloujifar, Tong Wu, Ruoxi Jia, Prateek Mittal
NeurIPS 2023 {PDF}
Recognized with Princeton's Early Career Graduate Award.

Data Banzhaf: A Robust Data Valuation Framework for Machine Learning
Jiachen T. Wang, Ruoxi Jia
AISTATS 2023 **(Oral Presentation, top 1.6% among submissions)** {PDF}
Featured in the course materials for Harvard CS236 (Topics in Computer Science and Economics).
Incorporated into open-source ML tools by AppliedAI Institute, a trustworthy AI consulting company.

LAVA: Data Valuation without Pre-Specified Learning Algorithms
Hoang Anh Just*, Feiyang Kang*, **Jiachen T. Wang**, Yi Zeng, Myeongseob Ko, Ming Jin, Ruoxi Jia
ICLR 2023 **(Spotlight Presentation, top 3% among submissions)** {PDF}
Incorporated into open-source ML tools by AppliedAI Institute, a trustworthy AI consulting company.

One-Round Active Learning through Data Utility Learning and Proxy Models
Jiachen T. Wang, Si Chen, Ruoxi Jia
TMLR {PDF}

Rényi Differential Privacy of Propose-Test-Release and Applications to Private and Robust ML
Jiachen T. Wang, Saeed Mahloujifar, Shouda Wang, Ruoxi Jia, Prateek Mittal
NeurIPS 2022 {PDF}

Improving Cooperative Game Theory-based Data Valuation via Data Utility Learning
Tianhao Wang, Yu Yang, Ruoxi Jia
ICLR 2022 Workshop on Socially Responsible Machine Learning {PDF}
Incorporated into open-source ML tools by AppliedAI Institute, a trustworthy AI consulting company.

Concurrent Composition of Differential Privacy
Salil Vadhan*, **Tianhao Wang***
TCC 2021, PPML 2021 Workshop, TPDP 2021 Workshop {PDF}
Invited to the Journal of Cryptology.

DPLUS: Boosting Utility of Differentially Private Deep Learning via Randomized Smoothing
Wenxiao Wang, **Tianhao Wang**, Lun Wang, Nanqing Luo, Pan Zhou, Dawn Song, Ruoxi Jia
PoPETS'21 {PDF}

Improving Robustness to Model Inversion Attacks via Mutual Information Regularization
Tianhao Wang, Yuheng Zhang, Ruoxi Jia
AAAI'21 {PDF}

A Principled Approach to Data Valuation for Federated Learning
Tianhao Wang, Johannes Rausch, Ce Zhang, Ruoxi Jia, Dawn Song
Book Chapter in Federated Learning: Privacy and Incentive (2020) {PDF}

---

*I have also **co-authored** papers across diverse research fields.*

A Sustainable Machine Learning Economy Needs Data Deals That Work for Generators
Ruoxi Jia, Luis Oala, Wenjie Xiong, Suqin Ge, **Jiachen T. Wang**, Feiyang Kang, Dawn Song
NeurIPS 2025 Position Paper Track

Boosting Alignment for Post-Unlearning Text-to-Image Generative Models
Myeongseob Ko, Henry Li, Zhun Wang, Jonathan Patsenker, **Jiachen T. Wang**, Qinbin Li, Ming Jin, Dawn Song, Ruoxi Jia
NeurIPS 2024 {PDF}

Language Models as Science Tutors
A. Chevalier, J. Geng, A. Wettig, H. Chen, S. Mizera, ..., **Jiachen T. Wang**, ..., Sanjeev Arora, Danqi Chen
ICML 2024 {PDF}

BaDExpert: Extracting Backdoor Functionality for Accurate Backdoor Input Detection
Tinghao Xie, Xiangyu Qi, Ping He, Yiming Li, **Jiachen T. Wang**, Prateek Mittal
ICLR 2024 {PDF}

Uncovering Adversarial Risks of Test-Time Adaptation
Tong Wu, Feiran Jia, Xiangyu Qi, **Jiachen T. Wang**, Vikash Sehwag, Saeed Mahloujifar, Prateek Mittal
ICML 2023 {PDF}

Towards a Proactive ML Approach for Detecting Backdoor Poison Samples
Xiangyu Qi, Tinghao Xie, **Jiachen T. Wang**, Tong Wu, Saeed Mahloujifar, Prateek Mittal
USENIX 2023 {PDF}

Just Rotate it: Deploying Backdoor Attacks via Rotation Transformation
Tong Wu, **Jiachen T. Wang**, Vikash Sehwag, Saeed Mahloujifar, Prateek Mittal
CCS Workshop on Artificial Intelligence and Security (AISec'22) {PDF}

RIGA: a Covert and Robust White-Box Watermarking of Deep Neural Networks
Tianhao Wang, Florian Kerschbaum
WWW'21 {PDF}

PREPRINTS & TECHNICAL WRITINGS

Can Small Training Runs Reliably Guide Data Curation? Rethinking Proxy-Model Practice
Jiachen T. Wang, Tong Wu, Kaifeng Lyu, James Zou, Dawn Song, Ruoxi Jia, Prateek Mittal
In submission, 2025 {PDF}
Impacted Google's data efforts for frontier AI development.

Does More Inference-Time Compute Really Help Robustness?
Tong Wu, Chong Xiang, **Jiachen T. Wang**, Weichen Yu, Chawin Sitawarin, Vikash Sehwag, Prateek Mittal
In submission, 2025 {PDF}

Effectively Controlling Reasoning Models through Thinking Intervention
Tong Wu, Chong Xiang, **Jiachen T. Wang**, G. Edward Suh, Prateek Mittal
In submission, 2025 {PDF}

An Economic Solution to Copyright Challenges of Generative AI
Jiachen T. Wang, Zhun Deng, Hiroaki Chiba-Okabe, Boaz Barak, Weijie Su
Technical Note, April 2024 {PDF}

A Note on "Efficient Task-Specific Data Valuation for Nearest Neighbor Algorithms"
Jiachen T. Wang, Ruoxi Jia
Technical Note, April 2023 {PDF}
Adopted in the technical design of PISTIS, a European funded project that aims to develop platforms for the sharing and trading of data assets.

A Note on "Towards Efficient Data Valuation Based on the Shapley Value"
Jiachen T. Wang, Ruoxi Jia
Technical Note, Feb 2023 {PDF}

SCHOLARSHIPS & AWARDS

ICLR 2025 Best Paper Honorable Mention *April 2025*
One of 6 papers recognized among 11,000+ submissions (for *"Data Shapley in One Training Run"*).

Apple Scholars in AIML PhD Fellowship *March 2025*
One of 21 recipients worldwide; full academic support from 2025-2027.

Yan Huo *94 Graduate Fellowship @ Princeton *Jan 2025*
One of 3 recipients department-wide; full academic support in Spring 2025.

Rising Stars in Data Science *Nov 2024*
One of 30 early-career researchers worldwide selected by UCSD, UChicago, and Stanford data science institutes.

Pramod Subramanyan Early Career Graduate Award @ Princeton *May 2023*
One of 2 second-year Ph.D. recipients department-wide for outstanding research and general exam performance.

Gordon Y. S. Wu Fellowship @ Princeton *Sept 2021*
For **the highest honor** that Princeton University can bestow on an incoming graduate student.

Winston and Diana Cherry Scholarship @ UWaterloo *May 2019*

For **the highest major average** at Spring Convocation with an Honors Statistics Major designation.

J.A. Brzozowski Scholarship @ UWaterloo *Dec 2018*
For outstanding achievement in courses focusing on **theoretical computer science**.

INVITED TALKS & TUTORIALS

Princeton ECE574 (Security and Privacy) Guest Lecture *Oct 2025*
Introduction to Differential Privacy

Apple Machine Learning Research Seminar *Aug 2025*
Can Small Training Runs Reliably Guide Data Curation? Rethinking Proxy-Model Practice

Google NYC Algorithms and Optimization Seminar *July 2025*
Data Shapley in One Training Run

Tsinghua University ML Foundations Seminar *Jan 2025*
Capturing the Temporal Dependence of Training Data Influence

NeurIPS 2024 Tutorial (2 hours, with Ruoxi Jia and Ludwig Schmidt) *Dec 2024*
Advancing Data Selection for Foundation Models: From Heuristics to Principled Methods

IDEAL Institute Workshop on Harmonious Human-AI Ecosystems *Nov 2024*
Capturing the Temporal Dependence of Training Data Influence

Virginia Tech ECE6514 (Trustworthy Machine Learning) Guest Lecture *Nov 2024*
Privacy-preserving Language Model Inference

Brandeis University Michtom School of Computer Science *Sept 2024*
Principled Data Attribution at Scale

Microsoft Research NYC Machine Learning Reading Group *Aug 2024*
An Influence-Preserving Approach to Continual Pretraining of LLMs

Google Privacy Seminar *Jan 2024*
TKNN-Shapley: A Linear-Time and Privacy-Friendly Approach to Data Valuation

Norwegian Computing Center Explaining AI Seminar *Jan 2024*
TKNN-Shapley: A Linear-Time and Privacy-Friendly Approach to Data Valuation

Google Privacy Seminar *May 2023*
A Randomized Approach to Tight Privacy Accounting

Boston-Area Data Privacy Seminar *Aug 2021*
Concurrent Composition of Differential Privacy

EXPERIENCES

Google Research **Remote**
Student Researcher *March 2025 - May 2026*
- Mentored by **Lin Chen**, **Mohammad Hossein Bateni**, and **Vahab Mirrokni** on *Gemini Data*.

Microsoft Research **New York City, NY**
Research Intern *June 2024 - Aug 2024*
- Mentored by **Huseyin A. Inan**, **Janardhan Kulkarni**, and **Ida Momennejad** on *continual pretraining of LLMs*.

Virginia Tech **Blacksburg, VA**
Research Assistant *June 2021 - Aug 2021*
- Mentored by **Prof. Ruoxi Jia** on *privacy-preserving machine learning* and *data-centric machine learning*.

Harvard University **Cambridge, MA**
Research Assistant *Aug 2020 - Aug 2021*
- Mentored by **Prof. Salil Vadhan** on the *theoretical foundation of differential privacy*.

University of California, Berkeley **Remote**
Visiting Student Scholar *May 2020 - Aug 2020*
- Mentored by **Prof. Dawn Song** and **Dr. Ruoxi Jia** on *data valuation for machine learning*.

University of Waterloo **Waterloo, ON**
Undergraduate Research Assistant *June 2018 - May 2019*
- Mentored by **Prof. Florian Kerschbaum** on *machine learning security.*

TEACHING

Information Security (COS432) @ Princeton *Spring 2024*
Graduate Teaching Assistant for **Prof. Maria Apostolaki**

Fairness and Privacy: Perspectives of Law and Probability (CS126) @ Harvard *Fall 2020*
Teaching Fellow for **Prof. Cynthia Dwork**

Statistics (STAT231) @ UWaterloo *Fall 2018*
Teaching Assistant

Algebra (MATH135) @ UWaterloo *Fall 2017*
Teaching Assistant

MENTORING

Mahavir Dabas (Junior PhD) *Since 2025*
- **Led to his first authorship on a conference paper submission.**

Weida Li (Junior PhD) *Since 2024*

Luxi He (Junior PhD) *Since 2023*

Tianji Yang (Undergraduate) *2023*
- **Led to his second authorship on an ICML 2024 oral paper.**

Tinghao Xie (Junior PhD) *Since 2023*
- **Led to his first authorship on an ICLR 2024 paper.**

Himanshu Jahagirdar (Master) *2023*
- **Led to his first authorship on a TMLR paper.**

SERVICE

Lead Organizer for Workshop on *Data Problems for Foundation Models* *ICLR 2025*
With **Ruoxi Jia**, **Pang Wei Koh**, **Dawn Song**, **Jerone Andrews**, Hoang Anh Just, Feiyang Kang.

Conference Paper Reviewer
- International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, 2024, 2025
- International Conference on Learning Representations (ICLR) 2023, 2024, 2025
- Conference on Neural Information Processing Systems (NeurIPS) 2023, 2024, 2025
- International Conference on Machine Learning (ICML) 2023, 2024, 2025
- Privacy Enhancing Technologies Symposium (PETS) 2021, 2022

Workshop Program Committee
- Data-centric Machine Learning ICML 2023, ICLR 2024, ICML 2024
- Privacy Regulation Protection in Machine Learning ICLR 2024

Research Mentor for AI4ALL Ignite Program *2024*
- Mentored $\approx$ **100 students** from underrepresented groups (Black, Indigenous, women, and non-binary) in developing AI/ML portfolio projects, providing both technical guidance and career advice.