

# Daogao Liu

Google Research  
liudaogao@gmail.com

## RESEARCH INTEREST

---

Machine Learning, Theoretical Computer Science  
Current Focus: Differential Privacy, Stochastic Optimization, and Algorithm Design  
Exploring: Large Language Models and Deep Learning

## EDUCATION

---

**University of Washington**, Seattle, US 2020 - 2024  
Ph.D. in Computer Science Advised by Yin Tat Lee  
Thesis: “Advancing Differentially Private Optimization: Efficiency, Utility, and Applications”  
  
**Tsinghua University**, Beijing, China 2016 - 2020  
B.S. in Physics and Mathematics Advised by Jian Li  
Thesis: “Stochastic Optimization: Approximation Algorithms on Metric Spaces”

## PROFESSIONAL EXPERIENCE

---

**Google Research**, Mountain View, US 2024 - present  
Visiting Postdoc Researcher, Working with the Algorithms team  
  
**Apple MLR**, Cupertino, US Summer, 2023/2024  
Research Intern, Working with Hilal Asi, Kunal Talwar and Vitaly Feldman  
  
**Google Brain**, Remote Spring 2023  
Research Intern, Working with Abhradeep Thakurta

## SELECTED LARGE-SCALE PROJECTS

---

- **Gemini — IMO Gold Medal (2025)** Data Contributor  
Official blog
- **VaultGemma — Pretraining with Differential Privacy Guarantees** Core Contributor  
Technical report, Official blog

## PUBLICATIONS

---

\* denotes equal contribution; most theoretical works are in (reverse) alphabetical order.

- [1] Syamantak Kumar, **Daogao Liu**, Kevin Tian, and Chutong Yang. Private Geometric Median in Nearly-Linear Time. *Accepted to NeurIPS 2025. Arxiv link*
- [2] Andrew Lowy\*, **Daogao Liu\***. Differentially Private Bilevel Optimization: Efficient Algorithms with Near-Optimal Rates. *Accepted to NeurIPS 2025. Arxiv link*
- [3] Da Yu, Edith Cohen, Badi Ghazi, Yangsibo Huang, Pritish Kamath, Ravi Kumar, **Daogao Liu**, and Chiyuan Zhang. Scaling Embedding Layers in Language Models. *Accepted to NeurIPS 2025. Arxiv link*
- [4] Weijia Shi, Akshita Bhagia, Kevin Farhat, Niklas Muennighoff, Jacob Morrison, Evan Pete Walsh, Dustin Schwenk, Shayne Longpre, Jake Poznanski, Allyson Ettinger, **Daogao Liu**, Margaret Li, Mike Lewis, Wen-tau Yih, Dirk Groeneveld + 8 more authors. FlexOLMo: Open Language Models for Flexible Data Use. *Accepted to NeurIPS 2025. Draft link*

- [5] **Daogao Liu**, Edith Cohen, Badih Ghazi, Peter Kairouz, Pritish Kamath, Alexander Knop, Ravi Kumar, Pasin Manurangsi, Adam Sealfon, Da Yu, and Chiyuan Zhang. Urania: Differentially Private Insights into AI Use. *Accepted to COLM 2025. Arxiv link*
- [6] Guy Kornowski, **Daogao Liu**, and Kunal Talwar. Improved Sample Complexity for Private Nonsmooth Nonconvex Optimization. *Accepted to ICML 2025. Arxiv link*
- [7] **Daogao Liu**, and Kunal Talwar. Adaptive Batch Size for Privately Finding Second-Order Stationary Points. *Accepted to ICLR 2025, Spotlight. Arxiv link*
- [8] Yangsibo Huang\*, **Daogao Liu\***, Lynn Chua, Badih Ghazi, Pritish Kamath, Ravi Kumar, Pasin Manurangsi, Milad Nasr, Amer Sinha, and Chiyuan Zhang. Unlearn and Burn: Adversarial Machine Unlearning Requests Destroy Model Accuracy. *Accepted to ICLR 2025. Arxiv link*
- [9] Weijia Shi, Jaechan Lee, Yangsibo Huang, Sadhika Malladi, Jieyu Zhao, Ari Holtzman, **Daogao Liu**, Luke Zettlemoyer, Noah A Smith, and Chiyuan Zhang. Muse: Machine unlearning six-way evaluation for language models. *Accepted to ICLR 2025. Arxiv link*
- [10] Hilal Asi, **Daogao Liu**, and Kevin Tian. Private Stochastic Convex Optimization with Heavy Tails: Near-Optimality from Simple Reductions. *Accepted to NeurIPS 2024. Arxiv link*
- [11] Hilal Asi, Tomer Koren, **Daogao Liu**, and Kunal Talwar. Private Online Learning via Lazy Algorithms. *Accepted to NeurIPS 2024. Arxiv link*
- [12] Andrew Lowy, **Daogao Liu**, and Hilal Asi. Faster Algorithms for User-Level Private Stochastic Convex Optimization. *Accepted to NeurIPS 2024. Arxiv link*
- [13] Lynn Chua, Badih Ghazi, Yangsibo Huang, Pritish Kamath, **Daogao Liu**, Pasin Manurangsi, Amer Sinha, and Chiyuan Zhang. Mind the Privacy Unit! User-Level Differential Privacy for Language Model Fine-Tuning. *Accepted to COLM 2024. Arxiv link*
- [14] Weijia Shi, Anirudh Ajith, Menthou Xia, Yangsibo Huang, **Daogao Liu**, Terra Blevin, Danqi Chen, and Luke Zettlemoyer. Detecting Pretraining Data from Large Language Models. *Accepted to ICLR 2024. Arxiv link*  
*Contribution: proposed the Min-K% Prob method (concept/design; implementation by collaborators).*
- [15] Gavin Brown, Krishnamurthy Dvijotham, Georgina Evans, **Daogao Liu**, Adam Smith, and Abhradeep Thakurta, Private gradient descent for linear regression: Tighter error bounds and instance-specific uncertainty estimation. *Accepted to ICML 2024. Arxiv link*
- [16] Hilal Asi, **Daogao Liu**, User-level Differentially Private Stochastic Convex Optimization: Efficient Algorithms with Optimal Rates. *Accepted to AISTATS 2024. Arxiv link*
- [17] Weijia Shi, Anirudh Ajith, Mengzhou Xia, Yangsibo Huang, **Daogao Liu**, Terra Blevins, Danqi Chen, Luke Zettlemoyer, Detecting pretraining data from large language models. *Accepted to ICLR 2024. Arxiv link*
- [18] Arun Ganesh, **Daogao Liu**, Sewoong Oh, Abhradeep Thakurta. Private (Stochastic) Non-Convex Optimization Revisited: Second-Order Stationary Points and Excess Risks. *Accepted to NeurIPS 2023, Spotlight. Arxiv link*
- [19] Yair Carmon, Arun Jambulapati, Yujia Jin, Yin Tat Lee, **Daogao Liu**, Aaron Sidford, Kevin Tian. ReSQueueing Parallel and Private Stochastic Convex Optimization. *Accepted to FOCS 2023. Arxiv link*
- [20] Sivakanth Gopi, Yin Tat Lee, **Daogao Liu**, Ruoqi Shen, Kevin tian. Algorithmic Aspects of the Log-Laplace Transform and a Non-Euclidean Proximal Sampler. *Accepted to COLT 2023. Arxiv link*
- [21] Hu Fu, Jiawei Li, **Daogao Liu**. Pandora Box Problem with Nonobligatory Inspection: Hardness and Improved Approximation Algorithms. *Accepted to STOC 2023. Arxiv link*

- [22] Sivakanth Gopi, Yin Tat Lee, **Daogao Liu**, Ruoqi Shen, Kevin Tian. Private Convex Optimization in General Norms. *Accepted to SODA 2023. Arxiv link*
- [23] Yaonan Jin, **Daogao Liu**, Zhao Song. Super-resolution in High Dimension: (Nearly) Linear Running Time and Sample Complexity. *Accepted to SODA2023. Arxiv link*
- [24] Ziqi Wang, Yuexin Wu, Frederick Liu, **Daogao Liu**, Le Hou, Hongkun Yu, Jing Li, Heng Ji. Augmentation with Projection: Towards an Effective and Efficient Data Augmentation Paradigm for Distillation. *Accepted to ICLR 2023. Arxiv link*
- [25] Xuechen Li\*, **Daogao Liu\***, Tatsunori Hashimoto, Huseyin A Inan, Janardhan Kulkarni, Yin Tat Lee, Abhradeep Guha Thakurta. When Does Differentially Private Learning Not Suffer in High Dimensions? *Accepted to NeurIPS 2022. Arxiv link*
- [26] Sivakanth Gopi, Yin Tat Lee, **Daogao Liu**. Private convex optimization via exponential mechanism. *Accepted to COLT 2022. Arxiv link*
- [27] **Daogao Liu**. Better Private Algorithms for Correlation Clustering. *Accepted to COLT 2022. Arxiv link*
- [28] Jian Li, **Daogao Liu**. Multi-token Markov Game with Switching Costs. *Accepted to SODA 2022. Arxiv link*
- [29] Janardhan Kulkarni, Yin Tat Lee, **Daogao Liu**. Private Non-smooth ERM and SCO in Subquadratic Steps. *Accepted by NeurIPS 2021, Spotlight. Arxiv link*
- [30] Haotian Jiang, Jian Li, **Daogao Liu**, Sahil Singla. Algorithms and Adaptivity Gaps for Stochastic k-TSP. *Accepted to ITCS 2020. Arxiv link*
- [31] **Daogao Liu**. More efficient Algorithms for Stochastic Diameter and Some Unapproximated Problems in Metric Space. *Accepted to COCOON 2019.*

## MANUSCRIPTS

---

- [1] Human-Level PII Detection with Large Language Models. *Manuscript in preparation*
- [2] Badih Ghazi, Ravi Kumar, **Daogao Liu**, and Pasin Manurangsi. Linear-Time User-Level DP-SCO via Robust Statistics. *Under submission. Arxiv link*
- [3] YinTat Lee, **Daogao Liu**, Zhou Lu. The Power of Sampling: Dimension-free Risk Bounds in Private ERM. *Arxiv link*

## SELECTED AWARDS

---

- |  |      |
|--|------|
| • Apple Scholars in AI/ML, Ph.D fellowship.  | 2024 |
| • Paul G. Allen School First-Year Ph.D Fellowship                                  | 2020 |
| • Ye Qisun Nomination Award (The highest honor of Physics undergraduate), Tsinghua | 2020 |
| • Scholarship for Comprehensive Excellence, Tsinghua University                    | 2019 |

## PROGRAM COMMITTEES SERVICE AND REVIEWING

---

- Conferences: SODA 2026/2025/2023, COLT 2025, ICML 2025, COLM 2025, ICLR 2025, AISTATS 2025, NeurIPS 2025/2024/2023, TPDP 2025/2024/2023, STOC 2024, ALT 2024, FOCS 2022/2021, RANDOM 2020
- Journals: Operations Research, Theoretical Computer Science, Information and Computation

## TEACHING EXPERIENCES

---

- CSE521 - Design and Analysis of Algorithms Fall 2024  
- Instructor: Thomas Rothvoss
- CSE 521: Design and Analysis of Algorithms Fall 2023  
– Instructor: Shayan Oveis Gharan