Nikhil Ghosh

Contact

Research Fellow
Center for Computational Mathematics
Flatiron Institute
162 5th Ave, New York, NY 10010

✉ nghosh@flatironinstitute.org
nikhilgsh.github.io
Google Scholar
github.com/nikhilgsh

Education

University of California, Berkeley 2019–2024
Ph.D. in Statistics. Advisors: Prof. Bin Yu and Prof. Song Mei.

California Institute of Technology 2015–2019
B.S. in Computer Science.

Employment

Flatiron Institute, Center for Computational Mathematics 2024–Present
Research Fellow, New York, NY

Microsoft Research 2023
Research Intern. Host: Greg Yang. Transferring optimal hyperparameters from small models to large models in finetuning settings.

Google Research 2022
Student Researcher. Host: Rina Panigrahy. Compute–performance tradeoffs for large transformers using conditional computation and external memory.

Awards & Honors

Erich L. Lehmann Award (Ph.D.), UC Berkeley Department of Statistics 2024

Princeton Theory of Deep Learning Summer School 2021

Student Travel Award, NeurIPS 2019

(⋆ denotes equal contribution; titles link to an online version)

Selected Publications

Optimization and Scaling

Nikhil Ghosh, Tetiana Parshakova, and Robert M. Gower. PoLoRA: A Preconditioned Orthogonalized LoRA Optimizer. *arXiv preprint (2026)*.

Nikhil Ghosh, Denny Wu, and Alberto Bietti. Understanding the Mechanisms of Fast Hyperparameter Transfer. *ICLR (2026)*.

Soufiane Hayou, **Nikhil Ghosh**, and Bin Yu. PLoP: Precise LoRA Placement for Efficient Finetuning of Large Models. *ICLR (2026)*.

Garrett Wen, Alberto Bietti, **Nikhil Ghosh**, Theodor Misiakiewicz, and Denny Wu. Fast Learning Rate Transfer for Gradient Descent in Sketched Linear Regression. *HiLD at ICML (2026) (Spotlight)*.

Soufiane Hayou, **Nikhil Ghosh**, and Bin Yu. The Impact of Initialization on LoRA Finetuning Dynamics. *NeurIPS (2024)*.

Soufiane Hayou⋆, **Nikhil Ghosh**⋆, and Bin Yu. LoRA+: Efficient Low Rank Adaptation of Large Models. *ICML (2024)*. Integrated into HuggingFace PEFT and LLaMA-Factory.

Efficient Architectures: Mixture-of-Experts and Sparsity

Zeliang Zhang, **Nikhil Ghosh**, Jiani Liu, Bin Yu, and Xiaodong Liu. Does a Global Perspective Help Prune Sparse MoEs Elegantly? *arXiv preprint (2026)*.

Cenk Baykal, Dylan Cutler, Nishanth Dikkala, **Nikhil Ghosh**, Rina Panigrahy, and Xin Wang. Alternating Updates for Efficient Transformers. *NeurIPS (2023) (Spotlight)*.

Nishanth Dikkala, **Nikhil Ghosh**, Raghu Meka, Rina Panigrahy, Nikhil Vyas, and Xin Wang. On the Benefits of Learning to Route in Mixture-of-Experts Models. *EMNLP (2023)*.

Foundations of Deep Learning

James B. Simon, Daniel Kunin, …, **Nikhil Ghosh**, …, Joseph Turnbull. There Will Be a Scientific Theory of Deep Learning. *arXiv preprint (2026)*.

Nikhil Ghosh, Spencer Frei, Wooseok Ha, and Bin Yu. The Effect of SGD Batch Size on Autoencoder Learning: Sparsity, Sharpness, and Feature Learning. *JMLR 26(49) (2025)*.

James B. Simon, Dhruva Karkada, **Nikhil Ghosh**, and Mikhail Belkin. More is Better in Modern Machine Learning: When Infinite Overparameterization is Optimal and Overfitting is Obligatory. *ICLR (2024)*.

Nikhil Ghosh and Mikhail Belkin. A Universal Trade-off Between the Model Size, Test Loss, and Training Loss of Linear Predictors. *SIMODS (2023)*.

Gal Kaplun*, **Nikhil Ghosh***, Saurabh Garg, Preetum Nakkiran, and Boaz Barak. Deconstructing Distributions: A Pointwise Framework of Learning. *ICLR (2023)*.

Nikhil Ghosh, Song Mei, and Bin Yu. The Three Stages of Learning Dynamics in High-Dimensional Kernel Methods. *ICLR (2022)*.

Service & Teaching

Short Courses Theory of Scaling in Modern Deep Learning, half-day course with Soufiane Hayou, STAI-X 2026, Harvard University.

Teaching Graduate Student Instructor for Concepts of Probability (Stat 134) at UC Berkeley (2020); Teaching Assistant for Introduction to Algorithms (CS 38, 2018) and Undergraduate Game Theory (PS/Ec 172, 2016) at Caltech.

Invited Talks Universal Tradeoffs for Linear Predictors at MoDL (2023), CMStatistics (2022), DeepMath (2022), and the Google Algorithms Reading Group (2022); Three Stages of Kernel Dynamics at TOPML (2022) and MoDL (2022); guest lectures on double descent, UC San Diego (2023).

Reviewing NeurIPS (2023), JMLR (2023), COLT (2022), ICML (2022).

Mentorship UC Berkeley Directed Reading Program (2019), randomized algorithms.