Avrajit Ghosh

Homepage | Google Scholar | Email: ghoshavr@berkeley.edu | Phone: +1-5173027289 | Addr: Berkeley, CA, 94704.

Research Focus

Deep Learning Theory, Generalization, Optimization, Large Language Models, AI Safety.

Employment

- **Simons Institute for the Theory of Computing, UC Berkeley** — Berkeley, USA
  Postdoctoral Fellow — *August 2025-2027*
  Host: Peter L. Bartlett, Bin Yu.

Education

- **Michigan State University** — Michigan, USA
  Ph.D. in Computational Mathematics, Science and Engineering — *Aug 2020 – July 2025*
  Thesis: Implicit Regularization of Hyperparameters in Deep Learning
  Award: Fitch H. Beach Award (Highest Engineering Honor for Graduate Research)
  Advisors: Rongrong Wang, Saiprasad Ravishankar.
  Committee: Sijia Liu, John Wright

Publications

Fields: Deep Learning Theory, Optimization (*ICLR, ICML, NeurIPS*).
★ denotes representative publications; * denotes equal contribution.

Preprints

1. **Ghosh, A.***, Beneventano, P.*, Walden, M., Tao, M., Bartlett, P., Yu, B. & Poggio, T., Harmonic stability: Stationary distributions of SGD beyond Lyapunov stability (Preprint). ★
2. Gatmiry, K.*, **Ghosh, A.***, Mirtaheri, P.*, Lee, J.D., Haghtalab, N., Abbe, E. & Bartlett, P., Learn Your Own Thoughts: Abstract Token Curriculum (arXiv-2026). ★
3. Gatmiry, K.*, Ng, T.*, Garg, A.*, **Ghosh, A.**, Abbe, E., Bartlett, P. & Yu, B., Blackwell-Guided Logit Search (BLS): Optimizing Multi-Turn Jailbreaks with Subliminal Steering (Preprint).
4. Heiss, J., Rogow-Patt, G., Shivnani, R., Chandran, N., Ge, J., **Ghosh, A.** & Yu, B., Controlling Representation Sharing in Multilingual Transformers: A Mechanistic Study (Preprint).

Refereed Publications

5. **Ghosh, A.**, Yu B., Warmuth M., Bartlett P., Hard labels sampled from Sparse Targets mislead Rotation Invariant Algorithms. *International Conference on Machine Learning (ICML-2026).* ★
6. **Ghosh, A.**, Cong, B., Yokota, R., Ravishankar, S., Wang, R., Tao, M., Khan, M. E., and Möllenhoff, T. Variational Learning Finds Flatter Solutions at the Edge of Stability. *Advances in Neural Information Processing Systems (NeurIPS 2025).* **(Spotlight, Top 3.1%)**
7. **Ghosh, A.***, Kwon, S.*, Wang, R., Ravishankar, S. & Qu, Q., Learning Dynamics of Deep Linear Networks Beyond the Edge of Stability. *International Conference on Learning Representations (ICLR-2025).*
8. **Ghosh, A.**, Zhang, X., Sun, K., Qu, Q., Ravishankar, S. & Wang, R., Optimal Eye Surgeon: Finding Image Priors through Sparse Generators at Initialization. *International Conference on Machine Learning (ICML-2024).*
9. **Ghosh, A.***, Lyu, H.*, Zhang, X., & Wang, R., Implicit Regularization in Heavy-ball Momentum Accelerated Stochastic Gradient Descent. *International Conference on Learning Representations (ICLR-2023).* **(Spotlight, Top 5%)**
10. **Ghosh, A.***, Alkhouri, I.*, Bell, E.*, Liang, S., Wang, R. & Ravishankar, S., Understanding Untrained Deep Models for Inverse Problems: Algorithms and Theory. *IEEE Signal Processing Magazine, Special Issue on Mathematics of Deep Learning, 2025.*
11. **Ghosh, A.**, McCann, M. T., Mitchell, M. & Ravishankar, S., Learning Sparsity-Promoting Regularizers using Bilevel Optimization. *SIAM Journal on Imaging Sciences (SIIMS), vol. 17, no. 1, pp. 31–60, 2024.*
12. Zhang, X., **Ghosh, A.**, Liu, G. & Wang, R., Improving Generalization of Complex Models under Unbounded Loss Using PAC-Bayes Bounds. *Transactions on Machine Learning Research (TMLR), 2024.*

Selected Workshops/Short Conference

13. Dern, N.*, Mazioud, S.*, Heiss, J.*, **Ghosh, A.**, McDonald, C.J., Clara, G. & Yu, B., Tracking Training Phases in Compositional Learning with Task-Agnostic Measures. *Mechanistic Interpretability Workshop at the International Conference on Machine Learning (ICML-2026).*

14. **Ghosh, A.**, Wang, R., PAC-Bayes Generalization bounds for Score Based Diffusion Models. *Dynamics at the Frontiers of Optimization, Sampling, and Games Workshop at NeurIPS 2025.*

15. Liang, S., Bell, E., **Ghosh, A.**, and Ravishankar, S., Pruning Unrolled Networks (PUN) at Initialization for MRI Reconstruction Improves Generalization. *Proceedings of the 58th Asilomar Conference on Signals, Systems, and Computers (Asilomar-2024).*

16. Lyu, G., Afshari, M., Zhang, X., Xue, Z., **Ghosh, A.**, Bashyal, B., Wang, R., Johnsohn, K., Towards Understanding Task-agnostic Debiasing Through the Lenses of Intrinsic Bias and Forgetfulness. *The 62nd Annual Meeting of the Association for Computational Linguistics (ACL-2024).*

17. **Ghosh, A.**, Mccann, M. & Ravishankar, S., Bilevel Learning of L1 Regularizers with Closed-Form Gradients. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2022).*

Research Experience

- **Research Intern** — RIKEN Center for Advanced Intelligence Project
  Variational Learning at Edge of Stability — *Aug 2024 - Nov 2024*
  Advisor: *Mohammad Emtiyaz Khan, Thomas Möllenhoff*

- **Graduate Research Assistant** — Michigan State University
  Implicit Regularization of Gradient Descent in Deep Learning — *April 2022 - Present*
  Advisor: *Rongrong Wang, Saiprasad Ravishankar, Qing Qu*

- **Charpak Research Intern** — INRIA Sophia Antipolis
  Classification of Extracellular Matrix by Curvelet Transform — *May 2019 - August 2019*
  Advisor: *Laure Blanc-Féraud*

Selected Awards

- **Rising Stars in Statistics and Data Science, Cornell University (2026).**: "Highlights outstanding early-career scholars."
- **CPAL Rising Star Award (2026)**: "Highlights 15 exceptional junior researchers worldwide at the start of their career."
- **Fitch H. Beach Award (2025)**: "Recognizes the most outstanding graduate researcher within the College of Engineering."
- **Top Area Chair and Outstanding Reviewer (2023-26)**: NeurIPS-26 (AC); NeurIPS-23, ICLR-24, NeurIPS-25 (reviewer).

Conference Services

- **Area-Chair**: NeurIPS-2026, CPAL-2026,27.
- **Conference Reviewer**: ICML (2024,25,26), ICLR (2024,25,27), NeurIPS (2023,24,25), ICIP (2022), ICASSP (2023).
- **Journal Reviewer**: TMLR, T-PAMI, SIAM Journal on Imaging Sciences(1), IEEE-SPM.

Seminar Organization

- **Fall 2025 - Fall 2026**: Organizer for the ML-AI theory seminar at Simons Institute for the Theory of Computing.
- **Fall 2023 - Spring 2024**: Organizer for the Departmental CMSE Seminar series.
- **11/10/2023**: Organizer for the Third CMSE Data Science conference (DISC-2023).

Teaching and Advising

- **Spring 2026**: Guest Lecturer for ECE-550 (hosted by Qing Qu), University of Michigan.
- **Fall 2025**: Guest Lecturer for CMSE-890 (hosted by Saiprasad Ravishankar), Michigan State University.
- **Spring 2025**: Co-Instructor of CMSE-382 (Optimization Methods in Data Science)
- **Spring 2022**: Teaching Assistant of CMSE-821 (Numerical Methods for Differential Equations) and CMSE-823 (Numerical Linear Algebra).
- **Spring 2021**: Teaching Assistant of CMSE-201 (Introduction To Computational Modeling).
- **2022-25**: Mentored several undergraduate students at MSU as part of REU program.

Invited Talks/Tutorials and Conference Orals

- **Talk at AI-ML theory seminar at Simons Institute for the Theory of Computing:**: Talk on "Phase transition in SGD beyond Lyapunov stability."
- **Talk at ELLIS Institute Tübingen**: CPAL Rising Star talk on *"Hard labels sampled from sparse target mislead rotation-invariant algorithms"*, Spring 2026.
- **Talk at Simons Institute for the Theory of Computing**: Meet the Fellows, Fall 2025 talk.
- **Invited Talk at Math Machine Learning seminar, MPI MiS + UCLA**: To present talk on *"The effect of low rank and stochasticity on Gradient Descent at the Edge of Stability".*
- **Invited Talk at UCLA**: Presented *"Dynamics of Gradient Descent at Edge of Stability for Deep Linear Networks"* to Professor Deanna Needell's group.
- **CPAL-2024 Tutorial Talk**: Presented *"Recent Advances in Machine Learning for Image Reconstruction: From Sparse Modeling to Deep Networks"* at the Conference on Parsimony and Learning (CPAL) in January 2024.
- **Invited Talk at Georgia Institute of Technology**: Presented *"Implicit regularization in gradient descent"* to SLIM group in October, 2023. Hosted by Professor Felix J. Herrmann.
- **ICLR-2023 Oral Talk**: Presented *"Implicit regularization of momentum-SGD"* as an oral talk at International Conference on Learning Representations 2023, held in Kigali, Rwanda.
- **ISBI-2023 Tutorial Talk**: Presented *"Recent Advances in Machine Learning for Image Reconstruction: Sparse Models to Deep Networks"* at the IEEE International Symposium on Biomedical Imaging (ISBI) on April 19, 2023.