Depen Morwani

+1 (857)928-6485 | depenmorwani@gmail.com | depenm.github.io

RESEARCH INTERESTS

Deep Learning Optimization, Science of Deep Learning

EDUCATION

PhD, *Computer Science* — Aug 2022 - May 2026 (Expected)
Harvard University, USA — GPA: 4.0/4.0

Masters with Thesis, *Computer Science* — Jul 2019 - May 2021
IIT Madras, India — GPA: 9.82/10

Bachelors with Honors, *Computer Science* — Jul 2012 - May 2016
IIT Bombay, India — GPA: 8.74/10

PUBLICATIONS

- **The Potential of Second-Order Optimization for LLMs: A Study with Full Gauss-Newton** arXiv
  International Conference on Learning Representations (ICLR), 2026
  Natalie Abreu, Nikhil Vyas, Sham Kakade, Depen Morwani
- **Seesaw: Accelerating Training by Balancing Learning Rate and Batch Size Scheduling** arXiv
  International Conference on Learning Representations (ICLR), 2026
  Alexandru Meterez, Depen Morwani, Jingfeng Wu, Costin-Andrei Oncescu, Cengiz Pehlevan, Sham Kakade
- **SOAP: Improving and Stabilizing Shampoo using Adam** arXiv
  International Conference on Learning Representations (ICLR), 2025
  Nikhil Vyas, Depen Morwani, Rosie Zhao, Itai Shapira, David Brandfonbrener, Lucas Janson, Sham Kakade
- **A New Perspective on Shampoo's Preconditioner** arXiv
  International Conference on Learning Representations (ICLR), 2025
  Depen Morwani, Itai Shapira, Nikhil Vyas, Eran Malach, Sham Kakade, Lucas Janson
- **Deconstructing What Makes a Good Optimizer for Language Models** arXiv
  International Conference on Learning Representations (ICLR), 2025
  Rosie Zhao, Depen Morwani, David Brandfonbrener, Nikhil Vyas, Sham Kakade
- **How Does Critical Batch Size Scale in Pre-training?** arXiv
  International Conference on Learning Representations (ICLR), 2025
  Hanlin Zhang, Depen Morwani, Nikhil Vyas, Jingfeng Wu, Difan Zou, Udaya Ghai, Dean Foster, Sham Kakade
- **Feature emergence via margin maximization: case studies in algebraic tasks** arXiv
  Spotlight (5% acceptance rate) at International Conference on Learning Representations (ICLR), 2024
  Depen Morwani, Benjamin L. Edelman, Costin-Andrei Oncescu, Rosie Zhao, Sham Kakade
- **Beyond Implicit Bias: The Insignificance of SGD Noise in Online Learning** arXiv
  Spotlight (3.5% acceptance rate) at International Conference on Machine Learning (ICML), 2024
  Nikhil Vyas, Depen Morwani, Rosie Zhao, Gal Kaplun, Sham Kakade, Boaz Barak
- **Simplicity Bias in 1-Hidden Layer Neural Networks** arXiv
  Neural Information Processing Systems (NeurIPS), 2023
  Depen Morwani, Jatin Batra, Prateek Jain, Praneeth Netrapalli
- **Feature-Learning Networks Are Consistent Across Widths At Realistic Scales** arXiv
  Neural Information Processing Systems (NeurIPS), 2023

Nikhil Vyas, Alex Atanasov, Blake Bordelon, Depen Morwani, Sabarish Sainathan, Cengiz Pehlevan

- **Inductive Bias of Gradient Descent for Weight Normalized Smooth Homogeneous Neural Nets** arXiv
  Algorithmic Learning Theory (ALT), 2022
  Depen Morwani, Harish G. Ramaswamy

PREPRINTS

- **LOTION: Smoothing the Optimization Landscape for Quantized Training** arXiv
  Under submission at International Conference on Machine Learning (ICML), 2026
  Mujin Kwun, Depen Morwani, Chloe Huangyuan Su, Stephanie Gil, Nikhil Anand, Sham Kakade
- **Adam or Gauss-Newton? A Comparative Study In Terms of Basis Alignment and SGD Noise** arXiv
  Under submission at International Conference on Machine Learning (ICML), 2026
  Bingbin Liu, Rachit Bansal, Depen Morwani, Nikhil Vyas, David Alvarez-Melis, Sham M. Kakade
- **Connections between Schedule-Free Optimizers, AdEMAMix, and Accelerated SGD Variants** arXiv
  Under submission at International Conference on Machine Learning (ICML), 2026
  Depen Morwani, Nikhil Vyas, Hanlin Zhang, Sham Kakade
- **A Simplified Analysis of SGD for Linear Regression with Weight Averaging** arXiv
  Workshop Optimization for Machine Learning (OPT), 2025
  Alexandru Meterez, Depen Morwani, Costin-Andrei Oncescu, Jingfeng Wu, Cengiz Pehlevan, Sham Kakade
- **AdaMeM: Memory Efficient Momentum for Adafactor** openreview
  Workshop on Advancing Neural Network Training (WANT), 2024
  Nikhil Vyas, Depen Morwani, Sham M. Kakade

RESEARCH INTERNSHIP

Google DeepMind, London | *Research Intern* Apr 2025 - Sep 2025

- Developing a theoretical understanding of the interaction between layer alignment and progressive sharpening, and proposing a regularizer for mitigating sharpening.

Microsoft Research, NYC | *Research Intern* Jun 2024 - Aug 2024

- Understanding the interaction between learning rate decay, weight averaging and edge of stability.

Google AI, India | *Research Associate - Deep Learning* Jun 2021 - Jul 2022

- Rigorously defined and established the notion of **simplicity bias** for 1-hidden layer neural networks. arXiv
- Proposed and verified the efficacy of *orthogonal projection* training method for mitigating simplicity bias on real-world datasets.

IIT Madras, India | *Project Associate - Reinforcement Learning* Aug 2018 - Jun 2019

- Modeled the behavior of Taxi Driver as a sequential decision making problem, and quantified the value of demand information received by the driver by cleverly combining **Monte-Carlo sampling** and **Value Iteration**. report

INDUSTRY EXPERIENCE

Goldman Sachs, India | *Financial Analyst - Short Term Rates Desk* Jun 2016 - Jul 2018

- Responsible for **pricing and risk** calculation of various rates products such as Swaps and Futures.
- Improved yield curve fitting by modeling it as a **Bayesian Inference** problem with Gaussian process prior and optimized the implementation by expressing the posterior as a sum of basis functions.

Samsung Electronics, S. Korea | *Summer Intern - Machine Learning* May 2015 - Jul 2015

- Predicted upcoming failure in Base Station, which is an important component of mobile networks.
- Increased the precision of the **Random Forest** Algorithm from 70% to 90% at a recall of 60%, by designing better features.

TEACHING EXPERIENCE

- Fall 2023 Teaching Fellow for **Algorithms for Data Science** (CS224) at Harvard University
- Spring 2019 and Spring 2020 Co-Instructor and Teaching Assistant for **Big Data Laboratory**(CS4830) at IIT Madras
- Spring 2020 and Fall 2021 Teaching Assistant for **Pattern Recognition and Machine Learning**(CS5691) at IIT Madras
- Fall 2019 Teaching Assistant for **Non-Linear Optimization**(CS5020) at IIT Madras
- Spring 2019 Teaching Assistant for **Introduction to Data Analytics**(MS4610) at IIT Madras

ACHIEVEMENTS

- Organized the **High-dimensional Learning Dynamics** (HiLD) workshop at ICML 2024.
- Secured **5th** position in the **Predicting Generalization in Deep Learning** (PGDL) competition held at *Neural Information Processing Systems (NeurIPS) 2020* arXiv

INTERESTS AND SKILLS

- *Languages* - C++, Python
- *Frameworks* - Pytorch, JAX
- *Interests* - Optimization, Probability, Statistical Learning Theory