

Mo Zhou

Email: mozhou17@cs.washington.edu

Website: mozhou7.github.io

Research Interests

I am broadly interested in the **science of AI**: understanding the principles that explain the behavior of modern learning systems. My current work focuses on the **mathematical foundations** of these systems, particularly how **training dynamics, model structure, and data** shape learning in **large overparameterized models**. I study these questions in two complementary ways: by developing theory to analyze complex models, and by using simple models to isolate the underlying mechanisms.

Academic Experience

University of Washington, Seattle, WA

2024–present

IFDS Postdoctoral Scholar, Paul G. Allen School of Computer Science & Engineering

Advisors: Simon S. Du and Maryam Fazel

Education

Duke University, Durham, NC

2019–2024

Ph.D. in Computer Science

Thesis: Optimization Dynamics in Mildly Overparametrized Models

Advisor: Rong Ge

Peking University, Beijing, China

2015–2019

B.S. in Statistics

Elite Education Program on Applied Math and Statistics

Publications and Preprints

* denotes equal contribution or alphabetical author ordering.

Publications

- **Learning Orthogonal Multi-Index Models Beyond Small Initialization: Incremental Learning, Competitive Dynamics and Symmetry**
Mo Zhou, Weihang Xu, Simon S. Du, Maryam Fazel
Conference on Neural Information Processing Systems (NeurIPS), 2026
- **Is $\sqrt{d}$ Separation Necessary for Gradient EM to Learn Gaussian Mixtures in High Dimensions?**
Yiran Zhang, Mo Zhou, Weihang Xu, Maryam Fazel, Simon S. Du
Conference on Neural Information Processing Systems (NeurIPS), 2026
- **Memorizing Long-tail Data Can Help Generalization Through Composition**
Mo Zhou*, Haoyang Ma*, Rong Ge
International Conference on Learning Representations (ICLR), 2026
- **Convergence Dynamics of Over-Parameterized Score Matching for a Single Gaussian**
Yiran Zhang, Weihang Xu, Mo Zhou, Maryam Fazel, Simon S. Du
International Conference on Learning Representations (ICLR), 2026
- **How Does Gradient Descent Learn Features — A Local Analysis for Regularized Two-Layer Neural Networks**
Mo Zhou, Rong Ge
Conference on Neural Information Processing Systems (NeurIPS), 2024
- **Implicit Regularization Leads to Benign Overfitting for Sparse Linear Regression**
Mo Zhou, Rong Ge
International Conference on Machine Learning (ICML), 2023
- **Understanding Edge-of-Stability Training Dynamics with a Minimalist Example**
Xingyu Zhu*, Zixuan Wang*, Xiang Wang, Mo Zhou, Rong Ge
International Conference on Learning Representations (ICLR), 2023

- **Depth Separation with Multilayer Mean-Field Networks**
Yunwei Ren, **Mo Zhou**, Rong Ge
*International Conference on Learning Representations (ICLR), 2023, **Spotlight***
- **Plateau in Monotonic Linear Interpolation — A “Biased” View of Loss Landscape for Deep Networks**
Xiang Wang, Annie N. Wang, **Mo Zhou**, Rong Ge
International Conference on Learning Representations (ICLR), 2023
- **Understanding The Robustness of Self-supervised Learning Through Topic Modeling**
Zeping Luo*, Shiyu Wu*, Cindy Weng*, **Mo Zhou**, Rong Ge
International Conference on Learning Representations (ICLR), 2023
- **Understanding Deflation Process in Over-parametrized Tensor Decomposition**
Rong Ge*, Yunwei Ren*, Xiang Wang*, **Mo Zhou***
Conference on Neural Information Processing Systems (NeurIPS), 2021
- **A Local Convergence Theory for Mildly Over-Parameterized Two-Layer Neural Network**
Mo Zhou, Rong Ge, Chi Jin
Conference on Learning Theory (COLT), 2021
- **Towards Understanding the Importance of Shortcut Connections in Residual Networks**
Tianyi Liu*, Minshuo Chen*, **Mo Zhou**, Simon S. Du, Enlu Zhou, Tuo Zhao
Conference on Neural Information Processing Systems (NeurIPS), 2019
- **Towards Understanding the Importance of Noise in Training Neural Networks**
Mo Zhou*, Tianyi Liu*, Yan Li, Dachao Lin, Enlu Zhou, Tuo Zhao
*International Conference on Machine Learning (ICML), 2019, **Oral***

Preprints / In Review

- **Global Convergence of Gradient EM for Over-Parameterized Gaussian Mixtures**
Mo Zhou*, Weihang Xu*, Maryam Fazel, Simon S. Du
Preprint, arXiv:2506.06584, 2025

Workshop Papers

- **Multi-head CLIP: Improving CLIP with Diverse Representations and Flat Minima**
Mo Zhou, Xiong Zhou, Li Erran Li, Stefano Ermon, Rong Ge
NeurIPS OPT Workshop, 2023

Industry and Visiting Experience

- Stanford University**, Stanford, CA – *Visiting Student Researcher* *Summer 2023*
Advisor: Tengyu Ma
 Studied mean-field analysis of multi-index models.
- AWS AI, Amazon**, Santa Clara, CA – *Applied Scientist Intern* *Summer 2022*
Advisors: Xiong Zhou and Erran Li
 Used optimization-inspired methods to analyze and improve CLIP models, leading to a paper at NeurIPS 2023 OPT Workshop.
- Georgia Institute of Technology**, Atlanta, GA – *Research Intern* *Summer 2018*
Advisor: Tuo Zhao
 Studied optimization-based theory of deep learning, leading to papers at ICML 2019 and NeurIPS 2019.

Research Mentoring

- University of Washington**, Seattle, WA *2025–2026*
Mentee: Yiran Zhang (undergrad at Tsinghua University; visiting intern in Simon Du’s group at UW)
 Mentored research on learning dynamics for EM and related models, leading to papers at ICLR 2026 and NeurIPS 2026.
- Duke University**, Durham, NC *2023–2024*
Mentee: Shiyu Wu (undergrad at Duke)
 Mentored a project on grokking that became the student’s senior thesis and received Highest Distinction.
- Duke University**, Durham, NC *Summer 2021*
Mentees: Zeping Luo, Cindy Weng, and Shiyu Wu (undergrads at Duke; through CS Plus Program)
 Mentored a project on self-supervised learning. The work resulted in a paper at ICLR 2023.

Honors and Awards

ICML Gold Reviewer	2026
ICLR Notable Reviewer	2025
NeurIPS Travel Award	2019
Houston BAA Scholarship, Peking University	2018
May 4th Scholarship, Peking University	2016, 2017
Academic Excellence Award, Peking University	2016, 2017, 2018

Talks

Learning Orthogonal Multi-Index Models Beyond Small Initialization: Incremental Learning, Competitive Dynamics and Symmetry

IFDS Research Expo, University of Wisconsin-Madison Aug. 2026

Global Convergence of Gradient EM for Over-Parameterized Gaussian Mixtures

IFDS Seminar, University of Washington Oct. 2025

Biennial Meeting of the Pacific Northwest Section of SIAM (PNW SIAM), Seattle Oct. 2025

International Conference on Continuous Optimization (ICCOPT) 2025, Los Angeles Jul. 2025

How Does Gradient Descent Learn Features — A Local Analysis for Regularized Two-Layer Neural Networks

Junior Theorists Workshop 2024 and NSF TRIPODS Workshop, TTIC Dec. 2024

IFDS Seminar, University of Washington Oct. 2024

Implicit Regularization Leads to Benign Overfitting for Sparse Linear Regression

INFORMS Annual Meeting, Phoenix Oct. 2023

Algorithm Seminar, Duke University Mar. 2023

A Local Convergence Theory for Mildly Over-Parameterized Two-Layer Neural Network

Theory of Overparameterized Machine Learning (TOPML), Online Apr. 2021

NSF-Simons Journal Club, Online Feb. 2021

Academic Service

Organizer

- IFDS Seminar Co-organizer Fall 2025–Spring 2026

Teaching

- Teaching Assistant, *CPS590.04 Machine Learning Algorithms*, Duke University Spring 2021
- Teaching Assistant, *CPS330 Design and Analysis of Algorithms*, Duke University Fall 2020
- Teaching Assistant, *CPS330 Design and Analysis of Algorithms*, Duke University Spring 2020

Conference

- Area Chair: ICLR (2027)
- Reviewer: ICML, ICLR, NeurIPS, COLT, STOC, AAAI, AISTATS, CVPR, ALT

Journal

- Reviewer: JMLR, Mathematical Programming, TPAMI, JASA, TMLR