

Junhong Shen

5000 Forbes Ave, Pittsburgh, PA, 15213 | ☎ (+1)310-689-6721 | ✉ junhongs@andrew.cmu.edu

🏠 [Website](#) | 💻 [GitHub](#) | 🎓 [Scholar](#)

Research Interests

Developing efficient, data-centric ML techniques for real-world ML applications

Education

Carnegie Mellon University

Pittsburgh, PA

PH.D. IN MACHINE LEARNING, ADVISOR: [AMEET TALWALKAR](#) (QPA: 4.11/4.33)

Sep. 2021 - May. 2026 (expected)

- Thesis Topic: Navigating Through Heterogeneous Data: Building AI Systems for Diverse Data Types, Domains, and Complexities
- Committee: Ameet Talwalkar (CMU/Datadog), Ruslan Salakhutdinov (CMU/Meta), Aviral Kumar (CMU/DeepMind), Ludwig Schmidt (Stanford/Anthropic), Alexander Toshev (Apple)

University of California, Los Angeles

Los Angeles, CA

B.S. IN MATHEMATICS OF COMPUTATION (GPA: 4.0/4.0)

Sep. 2017 - June 2021

- Daus Prize: Awarded to only 5 top undergraduate students in the mathematics department

Publications

PREPRINT

Thinking vs. Doing: Agents that Reason by Scaling Test-Time Interaction [\[paper\]](#)[\[code\]](#)[\[website\]](#)

Junhong Shen*, Hao Bai*, Lunjun Zhang, Yifei Zhou, Amrith Setlur, Shengbang Tong, Diego Caples, Nan Jiang, Tong Zhang, Ameet Talwalkar, Aviral Kumar

CAT: Content-Adaptive Image Tokenization [\[paper\]](#)

Junhong Shen, Kushal Tirumala, Michihiro Yasunaga, Ishan Misra, Luke Zettlemoyer, Lili Yu*, Chunting Zhou*

CodePDE: Benchmarking LLMs' Abilities to Solve PDEs through Code Generation [\[paper\]](#)[\[code\]](#)

Shanda Li, Tanya Marwah, **Junhong Shen**, Weiwei Sun, Andrej Risteski, Yiming Yang, Ameet Talwalkar

L2G: Repurposing Language Models for Genomics Tasks [\[paper\]](#)

Wenduo Cheng, **Junhong Shen**, Mikhail Khodak, Jian Ma, Ameet Talwalkar

PEER-REVIEWED ARTICLES

Mixture-of-Mamba: Enhancing Multi-Modal State-Space Models with Modality-Aware Sparsity

ICLR 2025 Scalable Optimization for Efficient and Adaptive Foundation Models Workshop (Oral, top 8/96) [\[paper\]](#)[\[code\]](#)

Weixin Liang*, **Junhong Shen***, Genghan Zhang, Ning Dong, Luke Zettlemoyer, Lili Yu

ScribeAgent: Towards Specialized Web Agents Using Production-Scale Workflow Data

ICLR 2025 Foundation Models in the Wild Workshop [\[paper\]](#)[\[code\]](#)[\[blog\]](#)

Junhong Shen, Atishay Jain, Zedian Xiao, Ishan Amlekar, Mouad Hadji, Aaron Podolny, Ameet Talwalkar

Specialized Foundation Models Struggle to Beat Supervised Baselines

ICLR 2025 [\[paper\]](#)[\[code\]](#)

Zongzhe Xu, Ritvik Gupta, Wenduo Cheng, Alexander Shen, **Junhong Shen**, Ameet Talwalkar, Mikhail Khodak

UPS: Towards Foundation Models for PDE Solving via Cross-Modal Adaptation

TMLR 2024 & ICML 2024 AI4Science Workshop (Spotlight) [\[paper\]](#)[\[code\]](#)

Junhong Shen, Tanya Marwah, Ameet Talwalkar

Tag-LLM: Repurposing General-Purpose LLMs for Specialized Domains

ICML 2024 [\[paper\]](#)[\[code\]](#)

Junhong Shen, Neil Tenenholtz, James Brian Hall, David Alvarez-Melis, Nicolò Fusi

Cross-Modal Fine-Tuning: Align then Refine

ICML 2023 (Oral, top 158/6538) [\[paper\]](#)[\[code\]](#) [\[talk\]](#)[\[website\]](#)

Junhong Shen, Liam Li, Lucio M. Dery, Corey Staten, Mikhail Khodak, Graham Neubig, Ameet Talwalkar

Efficient Architecture Search for Diverse Tasks

NeurIPS 2022 [[paper](#)][[code](#)][[blog](#)]

Junhong Shen*, Mikhail Khodak*, Ameet Talwalkar

NAS-Bench-360: Benchmarking Neural Architecture Search on Diverse Tasks

NeurIPS 2022 Datasets and Benchmarks Track [[paper](#)][[website](#)][[blog](#)]

Renbo Tu*, Nicholas Roberts*, Mikhail Khodak, **Junhong Shen**, Frederic Sala, Ameet Talwalkar

AutoML Decathlon: Diverse Tasks, Modern Methods, and Efficiency at Scale

NeurIPS 2022 Competitions Track [[paper](#)][[website](#)]

Nicholas Roberts, ... 24 authors ..., **Junhong Shen**, Evan Sparks

Iterative Teacher-Aware Learning

NeurIPS 2021 [[paper](#)][[code](#)]

Luyao Yuan, Dongruo Zhou, **Junhong Shen**, Jingdong Gao, Jeffrey Chen, Quanquan Gu, Ying Nian Wu, Song-Chun Zhu

Theoretically Principled Deep RL Acceleration via Nearest Neighbor Function Approximation

AAAI 2021 [[paper](#)][[code](#)]

Junhong Shen, Lin F. Yang

Mathematical Reconstruction of Patient-Specific Vascular Networks Based on Clinical Images and Global Optimization

IEEE Access [[paper](#)][[code](#)][[talk](#)]

Junhong Shen, Abdul Hannan Faruqi, Yifan Jiang, Nima Maftoon

Emergence of Pragmatics from Referential Game between Theory of Mind Agents

NeurIPS 2019 Emergent Communication Workshop [[paper](#)][[code](#)]

Luyao Yuan, Zipeng Fu, Jingyue Shen, Lu Xu, **Junhong Shen**, Song-Chun Zhu

* Equal Contribution

Work Experience

Google DeepMind

STUDENT RESEARCHER (MENTOR: [ALIREZA FATHI](#), [DAVID ROSS](#), [CORDELIA SCHMID](#))

- Working on visual reasoning agents.

Mountain View, CA

May 2025 - Present

FAIR at Meta

RESEARCH INTERN (MENTOR: [CHUNTING ZHOU](#), [LILI YU](#), [LUKE ZETTMAYER](#))

- Worked on adaptive image tokenization.

Seattle, WA

May 2024 - Dec. 2024

[Scribe](#) AI/ML

SENIOR MACHINE LEARNING RESEARCHER

- Worked on LLM-based web agents.

Pittsburgh, PA

Feb. 2024 - May 2024

Microsoft Research

RESEARCH INTERN (MENTOR: [DAVID ALVAREZ-MELIS](#), [NICOLÒ FUSI](#))

- Worked on adapting LLMs to specialized domains (e.g., low-resource languages, protein sequences, chemical formulas) via prompt tuning.

Cambridge, MA

May 2023 - Aug. 2023

Determined AI, Hewlett Packard Enterprise

RESEARCH INTERN (MENTOR: [LIAM LI](#))

- Worked on fine-tuning LLMs and vision transformers for scientific modalities via distribution alignment.

Pittsburgh, PA

Jun. 2022 - Dec. 2022

[SenseTime](#) Face ID Research

PRODUCT MANAGER INTERN

- Worked on 3D-structured-light Face ID; participated in 5 software version releases, bug fixing, and testing.

Beijing, China

Jun. 2018 - Sep. 2018

Honors & Awards

- 2025 **Wilson Center, Pathways to AI Policy Program**, fellow
- 2024 **J.P. Morgan AI Ph.D. Fellowship**, awardee (accepted)
- 2024 **Bloomberg Data Science Ph.D. Fellowship**, awardee (declined)
- 2023 **CMU MLD Two Sigma PhD Fellowship Nomination**, one of only 2 PhD students nominated program-wide
- 2021 **UCLA Daus Prize**, top-5 undergraduate students in mathematics
- 2017 - 21 **UCLA Dean's Honors List**, awardee

Talks

- Aug. 2025 **Production-Scale Workflow Data Empowers Specialized Web Agents**, Research Summit, Ai4 2025
- Apr. 2025 [LLM Meets Web Browsing](#), AIRe Lab @ CMU
- Mar. 2025 [LLM Meets Web Browsing](#), EFML Reading Group, Stanford
- Aug. 2024 [Repurposing LLMs for Long-Tail ML Applications](#), Research Summit, Ai4 2024
- Nov. 2023 [Machine Learning for Diverse Tasks](#), Guest Lecture, ML with Large Datasets, CMU 10605
- Nov. 2023 **Bridging LLMs and Long Tail ML Applications**, Catalyst Reading Group, CMU
- Mar. 2023 [Cross-Modal Fine-Tuning](#), AI4Science Talks
- Dec. 2022 [DASH: How to Search Over Convolutions](#), The AutoML Podcast
- Oct. 2022 **Tackling Diverse Tasks with Neural Architecture Search**, DLML Journal Club, Mayo Clinic

Professional Service

Co-organizer of [CMU Agent Workshop](#), 2024 & 2025

Co-organizer of [AutoML Decathlon](#), NeurIPS 2022 Competition Track

CMU MSML Admissions Committee Fall 2022

CMU MLD Open House Committee Spring 2024/2025

Conference Reviewer NeurIPS (2022/2023/2024), ICLR (2024/2025), ICML (2024/2025), AAAI (2025), CVPR (2025), ICCV (2025)

Teaching Assistant Deep Learning Systems (CMU 10714), ML in Practice (CMU 10718), Linear Algebra (UCLA Math 115A)

Research Experience

SAGE Lab

Pittsburgh, PA

ADVISOR: [AMEET TALWALKAR](#) (CMU)

June. 2021 - Present

- Ph.D. research on developing effective and efficient ML/AutoML tools for solving diverse tasks in practice.

Lin Yang's Group

Los Angeles, CA

ADVISOR: [LIN F. YANG](#) (UCLA)

Jan. 2020 - June 2021

- Studied sample-efficient reinforcement learning; proposed an algorithm for estimating the value functions using nearest neighbor function approximator; provided theoretical justification on the sample complexity.

Center for Vision, Cognition, Learning, and Autonomy (VCLA), UCLA

Los Angeles, CA

ADVISOR: [SONG-CHUN ZHU](#), [YING NIAN WU](#) (UCLA)

Jan. 2019 - June 2021

- Studied how theory of mind (ToM) can be integrated into various ML settings to improve algorithm efficiency.
- **Project 1: Multi-Agent Deep Reinforcement Learning with ToM**
 - Proposed an adaptive ToM algorithm in a referential game setting where the teacher and the student model each other's action likelihood while learning their own Q-functions; studied the emergent communication protocol between the agents.
- **Project 2: Efficient Learners in Iterative Machine Teaching**
 - Integrated ToM into machine teaching; improved teaching efficiency by having the learners model the teacher's training sample selection strategy with maximum likelihood estimation.
- **Project 3: Meta Machine Teaching**
 - Studied how meta-learning can be combined with machine teaching: the teacher monitors a group of students' learning processes, deduces their initial beliefs and learning models, and adapts its teaching scheme to each learner.

Computational Metastasis Lab, Fields Institute for Research in Mathematical Sciences

Toronto, Canada

ADVISOR: [NIMA MAFTOON](#) (UNIVERSITY OF WATERLOO)

Jul. 2019 - Sep. 2019

- Developed a vascular network reconstruction framework that uses (i) the main vessel skeletons segmented from clinical images and (ii) global constructive optimization algorithm to generate patient-specific cerebral vascular models; validated the geometric (lengths, radii) and hemodynamic (pressure, shear stress) properties of the models via histogram analysis and blood flow simulation.

Related Coursework

CS	Machine Learning, Artificial Intelligence, Information Theory, Computer Vision, Algorithms and Complexities, Advanced Programming, Operating Systems Principles, Deep Learning Systems, Computer Network Fundamentals
Math	Linear Algebra, Multivariable Calculus, Differential Equations, Discrete Mathematics, Probability Theory, Stochastic Processes, Mathematical Analysis, Complex Analysis, Optimization, Applied Numerical Methods, Game Theory
Stats	Statistical Modeling in Vision and Cognition, Computing and Inference in Vision and Cognition, Data Analysis and Regression

Professional Skills

Coding	Proficient: Python, C, C++, Bash, R Familiar: MATLAB, Java, Arduino
Tools	Git, TeX TeX, PyTorch, Tensorflow, Scikit-learn, OpenCV, OpenAI Gym, Google Cloud Platform, Docker, SolidWorks