

VAISHNAVH NAGARAJAN

www.vaishnavh.github.io
vaishnavh@google.com

Professional Career

- **Senior Research Scientist**, Google *May 2025 - Present*
- **Research Scientist**, Google *Oct 2021 - Apr 2025*
- **Ph.D. in Computer Science** *Aug 2015 - July 2021*
Carnegie Mellon University (CMU)
Advisor: *J. Zico Kolter*
Thesis: Explaining generalization in deep learning: progress and fundamental limits [\[arxiv\]](#)
- **Bachelors in Technology** *2011 - 2015*
Indian Institute of Technology (IIT) Madras
Advisor: *Balaraman Ravindran*
Thesis: KWIK Inverse Reinforcement Learning
GPA: 9.88/10.0 (Rank 2)

Preprints

- [1] **Deep sequence models memorize geometrically; it's unclear why** [\[arxiv\]](#)
Shahriar Noroozizadeh, Vaishnavh Nagarajan, Elan Rosenfeld, Sanjiv Kumar

Conference Publications

- [2] **Roll the dice and look before you leap: Going beyond the creative limits of next-token prediction** [\[arxiv\]](#)
Vaishnavh Nagarajan*, Chen Henry Wu*, Charles Ding, Aditi Raghunathan
International Conference on Machine Learning (ICML 2025)
Oral paper (1% acceptance) and winner of **Outstanding Paper Award** (6 out of 12k submissions)
- Also in ICLR 2025 Workshop on Workshop on Spurious Correlation and Shortcut Learning: Foundations and Solutions
- [3] **The pitfalls of next-token prediction** [\[arxiv\]](#)
Gregor Bachmann* and Vaishnavh Nagarajan*
International Conference on Machine Learning (ICML 2024)
- Also **oral** presentation in ICLR 2024 Workshop "How Far Are We From AGI?"
- [4] **Sharpness-Aware Minimization enhances feature quality via balanced learning** [\[openreview\]](#)
Jacob Mitchell Springer, Vaishnavh Nagarajan, Aditi Raghunathan
International Conference on Learning Representations 2024 (ICLR 2024)

- [5] **The cost of scaling down large language models: Fact recall deteriorates before in-context learning** [\[arxiv\]](#)
Tian Jin, Nolan Clement, Xin Dong, Vaishnavh Nagarajan, Michael Carbin, Jonathan Ragan-Kelley, Gintare Karolina Dziugaite
International Conference on Learning Representations 2024 (ICLR 2024)
- [6] **Think before you speak: Training language models with pause tokens** [\[arxiv\]](#)
Sachin Goyal, Ziwei Ji, Ankit Singh Rawat, Aditya Krishna Menon, Sanjiv Kumar, Vaishnavh Nagarajan.
International Conference on Learning Representations 2024 (ICLR 2024)
- Also in NeurIPS 2023 Workshop on Robustness of Few-shot and Zero-shot Learning in Foundation Models
- [7] **On student-teacher deviations in distillation: does it pay to disobey?** [\[arxiv\]](#)
Vaishnavh Nagarajan, Aditya Krishna Menon, Srinadh Bhojanapalli, Hossein Mobahi, Sanjiv Kumar.
In Advances in Neural Information Processing Systems 36 (NeurIPS 2023)
- [8] **ResMem: Learn what you can and memorize the rest** [\[arxiv\]](#)
Zitong Yang, Michal Lukasik, Vaishnavh Nagarajan, Zonglin Li, Ankit Singh Rawat, Manzil Zaheer, Aditya Krishna Menon, Sanjiv Kumar
In Advances in Neural Information Processing Systems 36 (NeurIPS 2023)
- [9] **Assessing generalization of SGD via disagreement.** [\[arxiv\]](#)
Yiding Jiang*, Vaishnavh Nagarajan*, Christina Baek and J. Zico Kolter
International Conference on Learning Representations 2022 (ICLR 2022)
Spotlight paper (5.2% acceptance)
- Also in ICML 2021 Workshop on Over-parameterization: Pitfalls and Opportunities
- [10] **Understanding the failure modes of out-of-distribution generalization.** [\[arxiv\]](#)
Vaishnavh Nagarajan, Anders Andreassen and Behnam Neyshabur
International Conference on Learning Representations 2021 (ICLR 2021)
- [11] **A learning theoretic perspective on local explainability.** [\[arxiv\]](#)
Jeffrey Li*, Vaishnavh Nagarajan*, Gregory Plumb and Ameet Talwalkar
International Conference on Learning Representations 2021 (ICLR 2021)
- [12] **Provably safe PAC-MDP exploration using analogies.** [\[arxiv\]](#)
Melrose Roderick, Vaishnavh Nagarajan and J. Zico Kolter
In Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS 2021)
- [13] **Uniform convergence may be unable to explain generalization in deep learning.** [\[arxiv\]](#)
Vaishnavh Nagarajan and J. Zico Kolter.
In Advances in Neural Information Processing Systems 32 (NeurIPS 2019)
Oral paper (0.55% acceptance) and winner of **The Outstanding New Directions Paper Award**
- *In Workshop on Understanding and Improving Generalization in Deep Learning. (ICML 2019; spotlight talk in workshop)*

- In IAS/Princeton Workshop on Theory of Deep Learning 2019 (**spotlight** talk)

- [14] **Deterministic PAC-Bayesian generalization bounds for deep networks via generalizing noise-resilience.** [\[arxiv\]](#)
Vaishnavh Nagarajan and J. Zico Kolter.
International Conference on Learning Representations 2019 (ICLR 2019)
- [15] **Revisiting adversarial risk.** [\[arxiv\]](#)
Arun Sai Suggala, Adarsh Prasad, Vaishnavh Nagarajan and Pradeep Ravikumar
In Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS 2019)
- [16] **Geriatix: Aging what you see and what you don't see. A file system aging approach for modern storage systems**
Saurabh Kadekodi, Vaishnavh Nagarajan, Greg Ganger and Garth Gibson
Proceedings of the 2018 USENIX Conference on Usenix Annual Technical Conference (ATC 2018)
- [17] **Gradient descent GAN optimization is locally stable** [\[arxiv\]](#)
Vaishnavh Nagarajan and J. Zico Kolter.
In Advances in Neural Information Processing Systems 30 (NeurIPS 2017)
Oral paper (1.2% acceptance)
- [18] **Lifelong learning in costly feature spaces.** [\[arxiv\]](#)
with Avrim Blum and Maria-Florina Balcan.
In Proceedings of the 28th International Conference in Algorithmic Learning Theory (ALT 2017)
- [19] **Learning-theoretic foundations of algorithm configuration for combinatorial partitioning problems.** [\[arxiv\]](#)
with Maria-Florina Balcan, Ellen Vitercik and Colin White.
In Proceedings of the 30th Annual Conference on Learning Theory (COLT 2017)
- [20] **Every team deserves a second chance: Identifying when things go wrong.** [\[PDF\]](#)
Vaishnavh Nagarajan*, Leandro S. Marcolino* and Milind Tambe.
In Proceedings of the 14th International Conference on Autonomous Agents and Multi-agent Systems (AAMAS 2015)

Journal Publications

- [21] **What do larger image classifiers memorise?** [\[arxiv\]](#)
Michal Lukasik, Vaishnavh Nagarajan, Ankit Singh Rawat, Aditya Krishna Menon, Sanjiv Kumar
Transactions on Machine Learning Research (TMLR 2024)

The following are journal versions of conference papers listed earlier.

- [22] **Lifelong learning in costly feature spaces.**
with Avrim Blum and Maria-Florina Balcan.
In Theoretical Computer Science (invited) (TCS 2019)

- [23] **Every team deserves a second chance: An extended study on predicting team performance.**
Leandro S. Marcolino, Aravind Lakshminarayanan, Vaishnavh Nagarajan and Milind Tambe.
In Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS 2016)

Workshop/Short papers

- [24] **Avoiding Spurious Correlations: Bridging Theory and Practice** [\[openreview\]](#)
Thao Nguyen, Vaishnavh Nagarajan, Hanie Sedghi, Behnam Neyshabur
NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications (2021)
- [25] **Theoretical insights into memorization in GANs.** [\[PDF\]](#)
Vaishnavh Nagarajan, Colin Raffel and Ian Goodfellow.
In Workshop on Integration of Deep Learning Theories (NeurIPS 2018)
- [26] **Generalization in deep learning: the role of distance from initialization.** [\[arxiv\]](#)
Vaishnavh Nagarajan and J. Zico Kolter.
In Workshop on Deep Learning: Bridging Theory and Practice (NeurIPS 2017)
spotlight talk in workshop)
- [27] **A reinforcement learning approach to online learning of decision trees.** [\[arxiv\]](#)
Abhinav Garlapati, Aditi Raghunathan, Vaishnavh Nagarajan and Balaraman Ravindran.
In Proceedings of the 12th European Workshop on Reinforcement Learning, International Conference on Machine Learning (EWRL-ICML 2015)
- [28] **KWIK inverse reinforcement learning.** [\[PDF\]](#)
Vaishnavh Nagarajan and Balaraman Ravindran.
The Multi-disciplinary Conference on Reinforcement Learning and Decision Making. (RLDM 2015)

Talks

- Going beyond the creative limits of next-token prediction
 - Amazon AGI Autonomy Reading Group (**invited**) Sep 2025
 - National University of Singapore AI Institute Talk (**invited**) Oct 2025
 - ICML 2025 Oral July 2025
 - Mila - Quebec AI Institute Seminar (**invited**) July 2025
 - Learning workshop (**invited**) April 2025
- The pitfalls of next-token prediction
 - ML at Flatiron Institute (**invited**) Sep 2025
 - CMU 15-789 Guest Lecture (**invited**) Fall 2024
 - CMU AI Lunch (**invited**) Fall 2024

- Simons Institute Workshop on Emerging Generalization Settings (**invited**) *Fall 2024*
 - Amazon Search Research Talk Series (**invited**) *Fall 2024*
 - Microsoft Research (**invited**) *Fall 2024*
 - Princeton Guest Lecture (**invited**) *Fall 2024*
 - NYU CILVR Seminar (**invited**) *Fall 2024*
 - ICLR 2024 Workshop “How Far Are We From AGI?” *May 2024*
- Assessing generalization via disagreement
 - ICLR 2021 spotlight talk *May 2021*
- Understanding the failure modes of out-of-distribution generalization.
 - IISA 2022 (**invited**) *Fall 2022*
 - CMU AI lunch *Mar 2021*
- Uniform convergence may be unable to explain generalization in deep learning.
 - CMU Lecture (**invited**) *Fall 2022*
 - Google Research (New York) Learning Theory (**invited**) *Oct 2020*
 - Center for Human Compatible AI, UC Berkeley (**invited**) *Aug 2020*
 - Google Brain (Mountain View) Deep Learning Phenomena (**invited**) *Jul 2020*
 - NeurIPS 2019 Oral presentation *Dec 2019*
 - CMU AI Lunch *Nov 2019*
 - IAS/Princeton University Workshop on Theory of Deep Learning: Where next? *Oct 2019*
 - ICML Workshop: Understanding and Improving Generalization in Deep Learning *Jun 2019*
- Generalization in deep learning: The role of distance from initialization
 - NeurIPS Workshop on Deep Learning: Bridging Theory and Practice *Dec 2017*
- Gradient Descent GAN optimization is locally stable.
 - NeurIPS 2017 Oral presentation *Dec 2017*
 - CMU AI lunch *Oct 2017*
- Lifelong learning in costly feature spaces.
 - ALT 2017 *Oct 2017*
- Learning the best algorithm for max-cut, clustering, and other partitioning problems.
 - Learning, Algorithm Design & Beyond Worst-Case Analysis, Simons Institute, Berkeley. (**invited**) *Nov 2016*
 - CMU Theory Lunch *Nov 2016*

- Member of Thesis Committee
 - Sachin Goyal, Machine Learning Department, Carnegie Mellon University, *Token-Efficient Scaling of Foundation Models: From Pretraining to Inference*, Principal Advisor: Zico Kolter.
- Session Chair: ICML 2025
- Area Chair:
 - ICML 2025
 - COLM 2025
 - NeurIPS 2025
 - NeurIPS 2023 R0-FoMo Workshop
- Reviewer:
 - ALT 2021
 - ICLR 2023, 2021 (**outstanding reviewer award**, top 10%);
 - NeurIPS 2024 (top 7% reviewer), 2023 (top 10% reviewer), 2021, 2020 (top 10% reviewer), 2019 (top 50% reviewer), 2018 (top 30% reviewer)
 - ICML 2024 (Expert Reviewer), 2023 (Expert Reviewer), 2022, 2021 (Expert reviewer), 2020, 2019 (**top 5%** reviewer)
 - COLT 2019
 - AISTATS 2023 (top 10% reviewer), 2019
 - UAI 2022
 - JMLR
 - Nature
 - Workshops: ICML PODS 2022, ICML OPPO 2021, ICLR 2023 ME-FoMo, NeurIPS 2023 DistShift
- Mentor at Learning Theory Alliance Workshop (**invited**) *Fall 2022 & 2023*
- Fatima Fellowship Mentor. *2022*
- Member of CMU Computer Science MS admissions committee. *Spring 2018*
- Representative of the Computer Science Department in the SCS4ALL PhD Committee, a student advisory council for the CMU School of Computer Science. *Fall 2017 - Fall 2018*
- Organized the Learning Theory Reading group in CMU. *Fall 2016*

Mentorship

- Shahriar Noorizadeh (PhD student at CMU as of 2025, hosted as intern at Google)
- Chen Henry Wu (PhD student at CMU as of 2025)

- Vansh Bansal (PhD student at UT Austin as of 2025)
- Aditya Gudibanda (Software Engineer at Google as of 2025, mentee through Google-internal scientific mentorship program)
- Peng-Yu Chen (Software Engineer at Google as of 2025, mentee through Google-internal scientific mentorship program)
- Gregor Bachmann (PhD student at ETH Zürich as of 2024)
- Sachin Goyal (PhD student at CMU as of 2024, hosted as intern at Google)
- Zitong Yang (PhD student at Stanford as of 2025)
- Jacob Springer (PhD student at CMU as of 2024)
- Yuri Galindo (mentee through Fatima Fellowship)
- Marcus Blake (Software Engineer at Google as of 2024, mentee through Learning Theory Alliance)
- Kimia Hamidieh (PhD student at MIT as of 2024)
- Nuredin Ali (PhD student at the University of Minnesota as of 2024)
- Thao Nguyen (PhD student at UW as of 2024)
- Yiding Jiang (PhD student at CMU as of 2024)
- Melrose Roderick (postdoc at Mila as of 2024)
- Jeffrey Li (PhD student at UW as of 2024)

Teaching

- Teaching assistant, 10-715: Advanced Introduction to Machine Learning *Fall 2016*
- Teaching assistant, 15-780: Graduate Artificial Intelligence *Spring 2018*

Internships

- **PhD Research Internship** *Summer 2020*
 Google X
Host: Behnam Neyshabur
 Theoretically explained when and why machine learning models fail to generalize under test-time distribution shifts.
- **PhD Research Internship** *Summer 2019*
 Bosch Center for AI
Host: David Reeb
 Developed generalization bounds for high-dimensional linear models that circumvent limitations of uniform convergence bounds.

- **PhD Research Internship**

Summer 2018

Google Brain

Host: Colin Raffel, Ian Goodfellow

Explained why Generative Adversarial Networks (GANs) counterintuitively do not memorize their training data. Explored metrics for measuring diversity of GAN samples and developed a theoretically-grounded technique for improving sample diversity.

- **Undergraduate Research Internship**

Summer 2014

University of Southern California (USC)

Advisor: Milind Tambe

Identified that a machine learning model can predict the success/failure of an artificial multi-agent team playing Computer Go.

- **Undergraduate Internship**

Summer 2013

[Report Bee](#)

Advisor: Madhavan Mukund, Chennai Mathematical Institute

Designed an index that quantifies learning experiences of schoolchildren. Implemented the model within Report Bee's web application.

Scholastic Achievements

- Among **national top 1%** candidates in national level olympiads (2011) in **five different fields**, namely, Informatics, Maths, Physics, Chemistry and Astronomy.
- Directly qualified for the Indian National Math Olympiad (INMO) 2011 based on outstanding performance in the Regional Math Olympiad 2010 and INMO 2010.
- One of the 35 students that qualified further for the national selection camp for International Chemistry Olympiad.
- Secured **All India Rank 70** (out of 0.5 million candidates) and State Rank 3 in IIT Joint Entrance Examination 2011, **All India Rank 56** (out of 1.1 million candidates) in All India Engineering Entrance Examination 2011 and **All India Rank 16** (out of 0.1 million candidates) in Indian Institute of Space Science and Technology Admission Test 2011

Honors and Awards

- Winner of **Outstanding Paper Award** at ICML 2025 (given to only 6 out of 12k submitted papers.)
- **Reviewer Award** at ICLR 2021 for being an outstanding reviewer.
- Winner of the **Outstanding New Directions Paper Award** at NeurIPS 2019 (given to only one out of $\sim 6500+$ submitted papers).
- Awarded the ACM-India/IARCS student grant to attend AAMAS 2015 in Istanbul, Turkey.

- One of ~ 30 Viterbi-India scholars selected by Viterbi School of Engineering (USC) and Indo-US Science and Technology Forum for a fully funded research internship in Summer 2014.
- Awarded the prestigious [KVPY](#) Fellowship 2009 by the **Government of India** to attract highly motivated students for pursuing a research career in science.
- Invited participant in the Council of Scientific and Industrial Research Programme on Youth for Leadership in Science 2009.

Other Activities

- Board member of CMU Indian Graduate Student Association (IGSA). *Dec 2015 - Dec 2018*
- National Service Scheme Volunteer involved in Scientific Toys & Assistive Technology. *2011-12*
- Taught basic maths to underprivileged primary school children in villages in India, in association with the NGO, *AID India*. *Dec 2011*
- Scribe for the students of [Vidya Sagar](#) (formerly, the Spastics Society of India). *2008-09*