Amrith Setlur

GHC 8013 CMU, 4902 Forbes Ave, Pittsburgh, PA 15213
✉ asetlur@andrew.cmu.edu • 🌐 ars22.github.io • Google Scholar

Education

Ph.D. in Machine Learning *Jan 2022 - May 2026*
Carnegie Mellon University, USA

M.S. in Language Technologies *Aug 2019 - June 2021*
Carnegie Mellon University, USA

B.Tech. with Honors in Computer Science and Engineering *July 2013 - August 2017*
National Institute of Technology, Trichy, India CGPA: 9.91/10.0 Institute Rank: 1 (**Gold Medalist**)

Industry Research

- **Meta Superintelligence Labs**, Menlo Park. *Research Intern.* *May 2025 – present*
  (Hosted by Sang Michael Xie & Paria Rashidinejad)
  Working on reallocating compute for scaling RL training of LLMs on hard problems by conditioning on very off-policy rollouts.
- **Google Research**, Mountain View. *Student Researcher.* *May 2024 – Dec 2024*
  (Hosted by Jacob Eisenstein, Jonathan Berant & Aviral Kumar)
  Worked on scaling automated process supervision (dense rewards) to improve test-time reasoning capabilities of LLMs.
- **Apple Machine Learning Research**, Cupertino. *Research Intern.* *May 2023 – Aug 2023*
  (Hosted by Kunal Talwar & Vitaly Feldman)
  Worked on using test-time personalization for improving private frequency estimation in real-world federated settings.
- **Amazon Research**, Bangalore. *Research Engineer.* *May 2017 – Mar 2019*
  Worked on bandit algorithms for improving ad recommendations.

Awards and Fellowships

Best Paper Award (*ICML Workshop on Exploration in AI Today*)	2025
JP Morgan AI PhD Fellowship	2024
SCS MLD Departmental Fellowship	2022
SCS LTI Departmental Fellowship	2020
President's Gold Medal (*NIT Trichy*)	2017
RECAL Alumni Award	2017
Institute Academic Prize (*NIT Trichy*)	2015, 2016, 2017
TATA CEDI Industrial Fellowship (*Best Undergraduate Thesis Award*)	2016
IASc Research Fellowship	2015
AIEEE Merit Scholarship (*JEE Mains All India Rank 196*)	2014
KVPY Fellowship	2012

References

1. **Prof. Virginia Smith**, Associate Professor of CS, Carnegie Mellon University. smithv@cmu.edu
2. **Prof. Sergey Levine**, Associate Professor of EECS, UC Berkeley. svlevine@eecs.berkeley.edu
3. **Prof. Yejin Choi**, Dieter Schwarz Professor of CS & HAI, Stanford University. yejinc@stanford.edu
4. **Prof. Aviral Kumar**, Assistant Professor of CS, Carnegie Mellon University. aviralku@andrew.cmu.edu
5. **Prof. Ruslan Salakhutdinov**, UPMC Professor of CS, Carnegie Mellon University. rsalakhu@andrew.cmu.edu

(Interfolio emails are available upon request.)

Relevant Publications And Pre-Prints

(* indicates Equal Contribution) .

Algorithms for Test-Time Adaptation

- **Thinking vs. Doing: Agents that Reason by Scaling Test-Time Interactions** [Paper] [Site]
  J. Shen*, H. Bai*, L. Zhang, Y. Zhou, **A. Setlur**, . . ., N. Jiang, T. Zhang, A. Talwalkar, A. Kumar
  Best Paper Award at the Multi-Modal Reasoning for Agentic Intelligence Workshop at ICCV 2025.
  Best Paper Honarable Mention at the Language Agents and World Models Workshop at NeurIPS 2025.
  Oral Presentation at the Scaling Environments for Agent Workshop at NeurIPS 2025.
  Neural Information Processing Systems (NeurIPS), 2025.
- **e3: Learning to Explore Enables Extrapolation of Test-Time Compute for LLMs** [Paper] [Site] [Blog]
  A. Setlur*, M. Yang*, C. Snell, J. Greer, I. Wu, V. Smith, M. Simchowitz, A. Kumar
  Oral Presentation at Test-Time Adaptation: Putting Updates to the Test (PUT) Workshop at ICML 2025.
  Oral Presentation at Long Context Foundation Models Workshop at ICML 2025.
  Best Paper Award at the Exploration in AI Today Workshop at ICML 2025.
- **RLAD: Training LLMs to Discover Abstractions for Solving Reasoning Problems** [Paper] [Site]
  Y. Qu, A. Singh, Y. Lee, **A. Setlur**, R. Salakhutdinov, C. Finn, A. Kumar
  Oral Presentation at RAM-2: Reasoning, Attention & Memory Workshop at COLM 2025.
- **Learning to Reason on Hard Problems with Privileged On-Policy Exploration** [Blog]
  Y. Qu*, **A. Setlur***, V. Smith, R. Salakhutdinov, A. Kumar
  Oral Presentation at MATH-AI Workshop NeurIPS 2025.
- **Scaling Test-Time Compute Without Verification or RL is Suboptimal** [Paper]
  A. Setlur, N. Rajaraman, S. Levine, A. Kumar
  Oral Presentation at the VerifAI Workshop at ICLR 2025.
  Spotlight presentation at the International Conference on Machine Learning (ICML), 2025.
- **Rewarding Progress: Scaling Automated Process Verifiers for LLM Reasoning** [Paper]
  A. Setlur*, C. Nagpal*, A. Fisch, X. Geng, J. Eisenstein, R. Agarwal, A. Agarwal, J. Berant, A. Kumar
  Spotlight presentation at the International Conference on Learning Representations (ICLR), 2025.
- **Optimizing Test-Time Compute via Meta Reinforcement Fine-Tuning** [Paper] [Site] [Blog]
  Y. Qu*, M. Yang*, **A. Setlur**, L. Tunstall, E. Beeching, R. Salakhutdinov, A. Kumar
  Oral Presentation at the FM-Wild Workshop at ICLR 2025.
  International Conference on Machine Learning (ICML), 2025.

Training Data for Test-Time Adaptation and Synthetic Data

- **What Do Learning Dynamics Reveal About Generalization in LLM Reasoning?** [Paper]
  K. Kang, **A. Setlur**, D. Ghosh, J. Steinhardt, S. Levine, A. Kumar
  International Conference on Machine Learning (ICML), 2025.
- **RL on Incorrect Synthetic Data Scales the Efficiency of LLM Math Reasoning by Eight-Fold** [Paper]
  A. Setlur, S. Garg, X. Geng, N. Garg, V. Smith, A. Kumar
  Neural Information Processing Systems (NeurIPS), 2024.
- **Deep Neural Networks Tend To Extrapolate Predictably** [Paper]
  K. Kang, **A. Setlur**, C. Tomlin, S. Levine
  International Conference on Learning Representations (ICLR), 2024.
- **Adversarial Unlearning: Reducing Confidence Along Adversarial Directions** [Paper]
  A. Setlur, B. Eysenbach, V. Smith, S. Levine
  Neural Information Processing Systems (NeurIPS), 2022.
- **Explaining The Efficacy of Counterfactually Augmented Data** [Paper]
  D. Kaushik, **A. Setlur**, E. Hovy, Z. Lipton
  International Conference on Learning Representations (ICLR), 2021.

Test-Time Adaptation for Robustness to Distribution Shifts

- **Prompting is a Double-Edged Sword: Improving Worst-Group Robustness of FMs** [Paper]
  A. Setlur, S. Garg, V. Smith, S. Levine

International Conference on Machine Learning (ICML), 2024.

- **Project and Probe: Sample-Efficient Adaptation by Interpolating Orthogonal Features** [Paper]
A. S. Chen*, Y. Lee*, **A. Setlur**, S. Levine, C. Finn
Spotlight presentation at the International Conference on Learning Representations (ICLR), 2024.
- **Complementary Benefits of Contrastive Learning and Self-Training Under Distribution Shift** [Paper]
S. Garg*, **A. Setlur***, Z. Lipton, S. Balakrishnan, V. Smith, A. Raghunathan
Neural Information Processing Systems (NeurIPS), 2023.
- **Contextual Reliability: When Different Features Matter in Different Contexts** [Paper]
G. Ghosal*, **A. Setlur***, D. Brown, A. Dragan, A. Raghunathan
International Conference on Machine Learning (ICML), 2023.
- **Bitrate-Constrained DRO: Beyond Worst Case Robustness To Unknown Group Shifts** [Paper]
A. Setlur, D. Dennis, B. Eysenbach, A. Raghunathan, C. Finn, V. Smith, S. Levine
International Conference on Learning Representations (ICLR), 2023.
- **Multitask Learning Can Improve Worst-Group Outcomes** [Paper]
A. Kulkarni*, L. Dery*, **A. Setlur**, A. Raghunathan, A. Talwalkar, G. Neubig
Transactions on Machine Learning Research (TMLR), 2023.
- **Two Sides of Meta-Learning Evaluation: In vs. Out of Distribution** [Paper]
A. Setlur*, O. Li*, V. Smith
Neural Information Processing Systems (NeurIPS), 2021.
- **Confidence-Based Model Selection: When to Take Shortcuts for Subpopulation Shifts** [Paper]
A. S. Chen, Y. Lee, **A. Setlur**, S. Levine, C. Finn

Test-Time Adaptation for Privacy Preservation with Better Utility

- **Exact Unlearning of Finetuning Data via Model Merging at Scale** [Paper]
K. Kuo, **A. Setlur**, K. Srinivas, A. Raghunathan, V. Smith
IEEE Secure and Trustworthy Machine Learning Conference (SaTML), 2026
- **Lower Bounds for Public-Private Learning under Distribution Shift** [Paper]
A. Setlur, P. Thaker, J. Ullman
- **Private and Personalized Frequency Estimation in a Federated Setting** [Paper]
A. Setlur, V. Feldman, K. Talwar
Neural Information Processing Systems (NeurIPS), 2024.
- **On the Benefits of Public Representations for Private Transfer Learning** [Paper]
P. Thaker, **A. Setlur**, V. Smith, Z. Steven Wu
Neural Information Processing Systems (NeurIPS), 2024.

Other Notable Works

- **Politeness Transfer: A Tag and Generate Approach** [Paper]
A. Madaan*, **A. Setlur***, . . . , G. Neubig, Y. Yang, R. Salakhutdinov, A. Black, S. Prabhumoye
Association for Computational Linguistics (ACL), 2020.
- **Nonlinear Independent Subspace Analysis for Learning Speech Representations** [Paper]
A. Setlur, B. Poczós, A. Black
Best Paper Finalist at Interspeech 2020.

Invited Talks and Research Blog Posts

Invited Talks

- **Extrapolating Test-Time Compute by Learning to Search** June 2025
Meta, Menlo Park (Post-Training Research Monthly)
- **Scaling Test-Time Compute Without Verification is Suboptimal** March 2025
Hosted by Andrea Zanette, Carnegie Mellon University
- **Test-Time Adaptation: Overview, Algorithms, and Challenges** February 2025
Hosted by Ameet Talwalkar, Carnegie Mellon University
- **Scaling Automated Process Verifiers for LLM Reasoning** November 2024

ServiceNow Research Seminar

- **Large-Scale Training on Suboptimal Synthetic Data** October 2024
  Guest Lecture in CMU 10605, ML with Large Datasets
- **Incorrect Synthetic Data Can Scale LLM Reasoning by 8×** July 2024
  Google DeepMind, Mountain View (Hosted by Aleksandra Faust)
- **Test-Time Training vs. Contrastive Learning for Distribution Shifts** May 2024
  Google Research Seminar on Foundation Models
- **Private and Personalized Frequency Estimation in a Federated Setting** March 2024
  Guest Lecture in CMU 10719, Federated and Collaborative Learning
- **When is Self-Training Reliable for Test-Time Adaptation?** October 2023
  Hosted by Sergey Levine, UC Berkeley
- **Test-Time Adaptation: Going Beyond Worst-Case Distribution Shifts** Aug 2023
  Apple Machine Learning Research
- **Reducing Confidence Along Adversarial Directions Improves Generalization** September 2022
  Hosted by Sergey Levine, UC Berkeley
- **Two Sides of Meta-Learning Evaluation: In vs. Out-of-Distribution** March 2022
  Hosted by Zachary Lipton, Carnegie Mellon University

Blog Posts

- **How to Explore to Scale RL Training of LLMs on Hard Problems?** November 2025
  Machine Learning Department Blog ***(5k+ views in <5 days)***
- **Sharpening or Discovery, RL or Meta RL?: How RL Improves LLM Reasoning** June 2025
  Notion Blog ***(30k+ views)***
- **Optimizing Test-Time Compute Involves Solving a Meta RL Problem** January 2025
  Machine Learning Department Blog ***(38k+ views)***

Research Mentoring

I have had the fortune of working with, advising, and mentoring amazing student collaborators.

PhD Students

- Katie Kang (UC Berkeley → Anthropic; Spring 2023 - Spring 2025)
- Annie Chen (Stanford University; Spring 2023 - Spring 2024)
- Yuxiao Qu (Carnegie Mellon University; since Fall 2024)
- Kevin Kuo (Carnegie Mellon University; Fall 2024 - Fall 2025)
- Junhong Shen (Carnegie Mellon University; Spring 2025 - Fall 2025)
- Anikait Singh (Stanford University; since Fall 2024)
- Lucio Dery (Carnegie Mellon University → Google DeepMind; Spring 2023 - Fall 2023)
- Lunjun Zhang (University of Toronto; Fall 2024 - Fall 2025)
- Jack Bai (University of Illinois Urbana-Champaign (UIUC); since Spring 2025)
- Ian Wu (Carnegie Mellon University; since Summer 2025)

Undergraduate & Master's Students

- Gaurav Ghosal (UC Berkeley CS Undergraduate → CMU PhD; Fall 2022 - Fall 2023)
- Matthew Yang (Master's in ML at CMU; since Fall 2024)
- Haoran Liu (Master's in ML at CMU; since Fall 2025)
- Raashi Mohan (CMU CS Undergraduate → Roblox; Spring 2024 - Fall 2024)
- Atharva Kulkarni (Master's in ML at LTI, CMU → USC PhD; Spring 2023 - Fall 2023)
- Ken Ziyu Liu (Master's in Robotics, CMU → Stanford PhD; Fall 2022 - Summer 2023)
- Abhinaya SB (NIT Trichy CS Undergraduate → PhD at NC State University; Spring 2022)

Professional Service

CMU Service

- CMU MLD PhD Admissions Committee 2023-2025
- CMU LTI MS Admissions Committee 2021

Workshop Organization

I was the lead co-organizer for the following workshops.

- ICLR 2025 Workshop on Scaling Self-Improving Foundation Models May 2025
  Singapore [Site]
- NeurIPS 2023 Workshop on Robustness of Few-Shot Learning in Foundation Models
  New Orleans [Site] December 2023

Conference and Journal Reviewing Duties

- Reviewer for NeurIPS (2021-2025) [**top 50% at NeurIPS 2024**], ICML (2022-2025), ICLR (2023-2025) [**notable reviewer at ICLR** 2025], NeurIPS Workshops Selection Committee (2024-2025), Transactions on Machine Learning (2024).
- Area Chair for the Next Generation of AI Safety Workshop ICML 2024.

Teaching Experience

- **Teaching Assistant** for Convex Optimization Spring 2023
  Course Number 10-725, Carnegie Mellon University
- **Teaching Assistant** for Federated and Collaborative Learning Fall 2024
  Course Number 10-719, Carnegie Mellon University
- **Teaching Assistant** for Combinatorics and Graph Theory Spring 2017
  Course Number CS 212, NIT Trichy
- **Teaching Assistant** for Data Structures and Algorithms Fall 2016
  Course Number CS 201, NIT Trichy
- **Teaching Assistant** for Advanced Calculus Spring 2016
  Course Number MA 102, NIT Trichy

Relevant Coursework

- **Graduate Courses:** 36709 Advanced Statistical Theory (A), 10716 Advanced Machine Learning Theory and Methods (A+), 10725 Convex Optimization (A+), 10708 Probabilistic Graphical Models (A+), 10701 Machine Learning (PhD) (A+), 11731 Machine Translation (A+), 11747 Neural Networks for NLP (A+), 11777 Multimodal Machine Learning (A).
- **Undergraduate Courses:** MA101 Advanced Calculus (A+), MA102 Graduate Linear Algebra (A+), MA204 Probability Theory (A+), CS064 Artificial Intelligence & Expert Systems (A+), CS065 Natural Language Processing (A+), CS201 Data Structures & Algorithms (A+), CS212 Combinatorics & Graph Theory (A+), CS203 Discrete Structures (A+), MA304 Operations Research (A+).

Outreach and Inclusion

- Mentor with the CMU Undergraduate Mentorship Program 2023
- Mentor with the Office of Alumni Relations, NIT Trichy, India 2018
- Mentor with the National Service Scheme at NIT Trichy, India 2014-2016