Amrith Setlur

asetlur@andrew.cmu.edu | ars22.github.io | ars22 | @setlur_amrith | Amrith Setlur

Research Overview

My research broadly focuses on addressing challenges in modern machine learning (ML) paradigms that remain unsolved after scaling data. As we exhaust internet data, my recent work has focused on enhancing generalization by developing new axes of sustained performance scaling through **self-training**, leveraging **synthetic data** with automated supervision, and optimizing models to use **test-time compute** via reinforcement learning. In the past, my work has spanned other challenges of modern ML like robustness to spurious correlations, out-of-distribution generalization, and data privacy leakage/memorization.

Education

Carnegie Mellon University — *Pittsburgh, PA*
PhD in Machine Learning — *Aug 2021 - Present*
CGPA: 4.05/4.0 Advisor: Prof. Virginia Smith

M.S. in Language Technologies — *Pittsburgh, PA* — *Aug 2019 - May 2021*
CGPA: 4.12/4.0 Advisors: Prof. Barnabás Póczos, Prof. Alan W Black

National Institute of Technology — *Trichy, India*
B.S. (Honors) in Computer Science and Engineering — *Aug 2013 - May 2017*
CGPA: 9.91/10.0; Institute Rank: 1 (**Gold Medalist**)

Selected Honors and Awards

- **JP Morgan AI PhD Fellowship 2024**
- **President's Gold Medal (Highest GPA in Undergrad)**
- **RECAL Alumni Award**
- **AIEEE Merit Scholarship**

Selected Publications

Self-Training, Synthetic Data & Test-Time Compute

Rewarding Progress: Scaling Automated Process Verifiers for LLM Reasoning
A. Setlur, C. Nagpal, A. Fisch, X. Geng, J. Eisenstein, R. Agarwal, A. Agarwal, J. Berant, A. Kumar; **ICLR 2025 (Spotlight)** [PDF]

Scaling Test-Time Compute Without Verification or RL is Suboptimal
A. Setlur, N. Rajaraman, S. Levine, A. Kumar; **VerifAI ICLR 2025 Workshop (Oral)** [PDF]

Optimizing Test-Time Compute via Meta Reinforcement Fine-Tuning
Y. Qu, M. Yang*, A. Setlur, L. Tunstall, E. Beeching, R. Salakhutdinov, A. Kumar*; **FM-Wild ICLR 2025 Workshop (Oral)** [PDF]

RL on Incorrect Synthetic Data Scales the Efficiency of LLM Math Reasoning by Eight-Fold
A. Setlur, S. Garg, X. Geng, N. Garg, V. Smith, A. Kumar; **NeurIPS 2024** [PDF]

Complementary Benefits of Contrastive Learning and Self-Training Under Distribution Shift
S. Garg, A. Setlur*, Z. Lipton, S. Balakrishnan, V. Smith, A. Raghunathan*; **NeurIPS 2023** [PDF]

Robustness to Spurious Correlations and Distribution Shifts

Prompting is a Double-Edged Sword: Improving Worst-Group Robustness of Foundation Models
A. Setlur, S. Garg, V. Smith, S. Levine; **ICML 2024** [PDF]

Contextual Reliability: When Different Features Matter in Different Contexts
G. Ghosal, A. Setlur*, D. Brown, A. Dragan, A. Raghunathan*; **ICML 2023** [PDF]

Bitrate-Constrained DRO: Beyond Worst Case Robustness To Unknown Group Shifts
A. Setlur, D. Dennis, B. Eysenbach, A. Raghunathan, C. Finn, V. Smith, S. Levine; **ICLR 2023** [PDF]

Two Sides of Meta-Learning Evaluation: In vs. Out of Distribution
A. Setlur, O. Li*, V. Smith*; **NeurIPS 2021** [PDF]

Memorization and Privacy Leakage

Private and Personalized Frequency Estimation in a Federated Setting
A. Setlur, V. Feldman, K. Talwar; **NeurIPS 2024** [PDF]

On the Benefits of Public Representations for Private Transfer Learning under Distribution Shift
P. Thaker, A. Setlur, V. Smith, Z. Steven Wu; **NeurIPS 2024** [PDF]

Adversarial Unlearning: Reducing Confidence Along Adversarial Directions
A. Setlur, B. Eysenbach, V. Smith, S. Levine; **NeurIPS 2022** [PDF]

Other Notable Works

Project and Probe: Sample-Efficient Domain Adaptation by Interpolating Orthogonal Features
Annie S. Chen, Yoonho Lee*, Amrith Setlur, Sergey Levine, Chelsea Finn*; **ICLR 2024 (Spotlight)** [PDF]

Explaining The Efficacy of Counterfactually Augmented Data
D. Kaushik, A. Setlur, E. Hovy, Z. Lipton; **ICLR 2021** [PDF]

Nonlinear Independent Subspace Analysis with Auxiliary Variables for Learning Speech Representations
A. Setlur, B. Poczós, A. Black; **Interspeech 2020 (Best Student Paper Finalist)** [PDF]

Politeness Transfer: A Tag and Generate Approach
A. Madaan, A. Setlur* , T. Parekh*, B. Póczos, G. Neubig, Y. Yang, R. Salakhutdinov, A. Black, S. Prabhumoye*; **ACL 2020** [PDF]

Notable Accolades

- **Graduate Fellowships:** JP Morgan AI Phd Fellowship 2024 and LTI Graduate Research Fellowship 2019-2021. IASc Research Fellowship 2015, where I worked on K-medians with outliers, with Prof. Naveen Garg at IIT Delhi.
- **Academic awards:** Recipient of three consecutive Institute Medals for the years 2014–15, 2015–16, 2016–17; President of India Gold Medal for securing the highest CGPA (9.91) across all departments; and the RECAL Alumni Award for "Best Academic Performance" and highest GPA in Computer Science Department. Secured All India Rank 196 at JEE Mains Examination (AIEEE) 2013 for which I was awarded AIEEE Merit Scholarship from HRD Ministry, Government of India (INR 140,000). I also won the CEDI TATA Industrial Grant of INR 40,000 for *Most Innovative Undergraduate Thesis* on extractive text-summarization.
- **ACM ICPC:** Represented the college at the ACM ICPC Regionals 2014 (Team Leader) and 2015 (Coach). Secured 1st positions at the Delta Algothon and the Algorithmic Coding Triathlon (Vortex) with 5000+ participants across the nation.

Industry Experience

Google Research (Hosted by **Jacob Eisenstein**, **Jonathan Berant** & **Aviral Kumar**) *San Francisco, USA*
Student Researcher *May 2024 - current*
Worked on scaling process reward models without human supervision for improving reasoning in large language models. Our dense rewards improved the compute efficiency of test-time search by $2\text{-}5\times$, and the efficiency of online sampling during reinforcement learning by $6\times$.

Apple ML Research (Hosted by **Kunal Talwar** & **Vitaly Feldman**) *Cupertino, USA*
Research Scientist Intern *May 2023 - Aug 2023*
Worked on private and personalized frequency estimation in a federated setting. Used Good-Turing estimation and clustering to develop algorithms that learn personalized histograms with provable user-privacy, showing practical gains on heterogeneous real-world datasets.

Amazon Research *Bengaluru, IND*
Machine Learning Engineer *Aug 2017 - May 2019*
Worked on the Ad Placement Optimization team, tackling online claim detection and personalized product recommendations. Developed ML algorithms for the "Frequently Bought Together" widget, boosting Sponsored Ads click-through rate by 0.4% on Amazon.com.

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+).

Teaching

- Teaching Assistant for 10725 Convex Optimization (Spring '23)
- Teaching Assistant for 10719 Federated and Collaborative Learning (Fall '23)

Service

- Reviewer for NeurIPS (21', 22', 23', 24'), ICML (22', 23', 24', 25'), ICLR (23', 24', 25').
- Co-organized NeurIPS 2023 Workshop on Robustness of few-shot and zero-shot learning in foundation models.
- Co-organizing ICLR 2025 Workshop on Scaling self-improving foundation models without human supervision.