Vinod Raman

vkraman@umich.edu
https://vinodkraman.github.io

Education

University of Michigan Ann Arbor, MI
PhD Statistics Sep. 2021 - Dec. 2025
Thesis Advisor: Ambuj Tewari
Research Areas: LLM Inference-time Methods, Synthetic Data Generation, Differential Privacy, Robustness

University of Michigan Ann Arbor, MI
BSE Computer Science, BSE Chemical Engineering Sep. 2015 - May 2020
Thesis Advisors: Mahdi Cheraghchi, Sindhu Kutty, Andrej Lenert
Research Areas: Submodular Optimization, Bandits, Thermophotovoltaics

Industry Experience

Google Research Ann Arbor, MI
Student Researcher Sep. - Nov. 2025
Hosts: Umar Syed

- Fine-tuning and filtering the outputs of LLMs for synthetic data generation.

Google Research New York City, NY
Research Intern June - Sep. 2025
Hosts: Matthew Joseph, Travis Dick, Umar Syed

- Designed and implemented new private algorithms for domain discovery, top-k selection, and submodular maximization over unbounded domains. Algorithms achieve state-of-the-art privacy-utility guarantees. Paper in submission at ICLR 2026.
- Implemented and ran an evaluation pipeline to benchmark the arithmetic capabilities of flagship LLMs (ChatGPT, Gemini, etc.). Paper in progress.
- Implemented a novel synthetic data generation pipeline involving fine-tuning and filtering the outputs of LLMs; used JAX to scale filtering method to process tens of thousands of synthetic examples.

Apple New York City, NY
Research Intern Feb. - June 2025
Host: Satyen Kale

- Designed and implemented adaptive allocation strategy for Best-of-N alignment for batch LLM inference. Paper in submission at ICLR 2026.
- Proposed efficient methods for sample-dependent mass estimation. Paper in progress.

Apple Cupertino, CA
Research Intern May - Aug. 2024
Host: Kunal Talwar

- Developed new differentially private algorithms achieving state-of-the-art privacy-utility tradeoffs for adversarial bandits and dynamic regret minimization.
- Research published in two papers at ICML 2025.

Technical Skills: Python, C++, Java; PyTorch, JAX, TensorFlow, HuggingFace

Awards & Fellowships

- ALT Outstanding Paper Award, 2025
- Apple Scholars in AI/ML PhD Fellowship, 2025
- Outstanding First-Year Ph.D. Student, 2022
- NSF Graduate Research Fellowship, 2022

Publications

(*) denotes equal contribution. $(\alpha\text{-}\beta)$ denotes alphabetical ordering

1. J. Li*, **V.Raman***, A. Tewari. Generation through the lens of learning theory.
   Conference on Learning Theory (COLT), 2025.
   https://arxiv.org/abs/2410.13714

2. A. Raman*, **V.Raman***. Generation from Noisy Examples.
   International Conference on Machine Learning (ICML), 2025.
   https://arxiv.org/abs/2501.04179

3. C. Peale*, **V.Raman***, O. Reingold*. Representative Language Generation.
   International Conference on Machine Learning (ICML), 2025.
   https://arxiv.org/abs/2505.21819

4. H. Asi*, **V.Raman***, A. Saha*. Tracking the Best Expert Privately.
   International Conference on Machine Learning (ICML), 2025.
   https://arxiv.org/abs/2503.09889

5. **V.Raman**, H. Asi, K. Talwar. Faster Rates for Private Adversarial Bandits.
   International Conference on Machine Learning (ICML), 2025.
   https://arxiv.org/abs/2505.21790

6. **V.Raman***, U.Subedi*, A.Tewari. The Complexity of Sequential Prediction in Dynamical Systems.
   Conference on Learning for Dynamics and Control (L4DC), 2025. Oral Presentation.
   https://arxiv.org/abs/2402.06614

7. **V.Raman***, U.Subedi*, A.Tewari. A Unified Theory of Supervised Online Learnability.
   Conference on Algorithmic Learning Theory (ALT), 2025. Outstanding Paper Award.
   https://arxiv.org/abs/2307.03816

8. **V.Raman***, U.Subedi*, A.Tewari. A Characterization of Multiouput Learnability.
   Journal of Machine Learning Research (JMLR), 2024.
   https://arxiv.org/abs/2301.02729

9. **V.Raman**, A.Tewari. Online Classification with Predictions.
   Conference on Neural Information Processing Systems (NeurIPS), 2024.
   https://arxiv.org/abs/2405.14066

10. S.Hanneke*, **V.Raman***, A. Shaeiri*, U.Subedi*. Multiclass Transductive Online Learning.
    Conference on Neural Information Processing Systems (NeurIPS), 2024. Spotlight.

11. **V.Raman***, U.Subedi*, A.Tewari. Smoothed Online Classification can be Harder than Batch Classification.
    Conference on Neural Information Processing Systems (NeurIPS), 2024.
    https://arxiv.org/pdf/2405.15424

12. **V.Raman***, U.Subedi*, A. Raman, A.Tewari. Apple Tasting: Combinatorial Dimensions and Minimax Rates.
    Conference on Learning Theory (COLT), 2024.
    https://arxiv.org/abs/2310.19064

13. **V.Raman***, U.Subedi*, A.Tewari. Online Learning with Set-Valued Feedback.
    Conference on Learning Theory (COLT), 2024.
    https://arxiv.org/abs/2306.06247

14. **V.Raman***, U.Subedi*, A.Tewari. Online Infinite-Dimensional Regression: Learning Linear Operators.
Conference on Algorithmic Learning Theory (ALT) 2024.
https://arxiv.org/abs/2309.06548

15. A.Raman, **V.Raman***, U.Subedi*, I.Mehalel*, A.Tewari. Multiclass Online Learnability under Bandit Feedback.
Conference on Algorithmic Learning Theory (ALT) 2024.
https://arxiv.org/abs/2308.04620

16. **V.Raman***, U.Subedi*, A.Tewari. On Proper Learnability between Average- and Worst-case Robustness.
Conference on Neural Information Processing Systems (NeurIPS) 2023.
https://arxiv.org/abs/2211.05656

17. **V.Raman***, U.Subedi*, A.Tewari. On the Learnability of Multilabel Ranking.
Conference on Neural Information Processing Systems (NeurIPS) 2023. Spotlight.
https://arxiv.org/abs/2304.03337

18. S.Hanneke*, S.Moran*, **V.Raman***, U.Subedi*, A.Tewari. Multiclass Online Learning and Uniform Convergence.
Conference on Learning Theory (COLT) 2023.
https://arxiv.org/abs/2303.17716

19. **V.Raman**, A.Tewari. Online Agnostic Multiclass Boosting.
Conference on Neural Information Processing Systems (NeurIPS) 2022.
https://arxiv.org/abs/2205.15113

20. **V.Raman**, T.Burger, A.Lenert. Design of thermophotovoltaics for tolerance of parasitic absorption.
Optics Express, 27(22):31757–31772, 2019.
https://doi.org/10.1364/OE.27.031757

In Submission

1. Y. Kalayci*, **V. Raman***, S. Dughmi. Optimal Stopping vs. Best-of-N for Inference-time Optimization.
In Submission, 2025.
https://arxiv.org/abs/2510.01394

2. S. Xie, **V. Raman**, S. Zhou. Transductive and Learning-Augmented Online Regression.
In Submission, 2025.

3. **V. Raman**, T. Dick, M. Joseph. Missing Mass for Differentially Private Domain Discovery.
In Submission, 2025.

4. **V. Raman**, H. Asi, S. Kale. AdaBoN: Adaptive Best-of-N Alignment.
In Submission, 2025.
https://arxiv.org/abs/2505.12050

5. S. Somerstep, **V. Raman***, U. Subedi*, Y. Sun. Learning to Choose or Choosing to Learn: Best-of-N vs. Supervised Fine-Tuning for Bit String Generation.
In Submission, 2025.
https://www.arxiv.org/abs/2505.17288

Preprints

1. (α-β)V. Feldman, S. Kale, **V. Raman**, K. Talwar, A. Tewari. Estimating the (Un)seen: Sample-dependent Mass Estimation. *Preprint*, 2025.

2. **V.Raman***, D.Zhang*, Y.Jung, A.Tewari. Online Boosting for Multilabel Ranking with Top-k Feedback.
Preprint, 2020.
https://arxiv.org/abs/1910.10937

Invited Talks

1. Optimal Stopping vs. Best-of-N for Inference-time Optimization. *Google DeepMind Tech Talk*, 2025.
2. A Unified Theory of Supervised Online Learning. *ALT*, 2025.
3. Generation through the lens of learning theory. *Apple MLR Reading Group*, 2025.
4. Generation through the lens of learning theory. *NEU CS Theory Seminar*, 2024.
5. Trichotomies in Online Learnability. *Apple MLR Reading Group*, 2024
6. Multiclass Online Learning and Uniform Convergence. *UMich EECS Theory Seminar*, 2024.
7. On Classification-Calibration of Gamma-Phi Losses. *COLT*, 2023.