

# Amey Pasarkar

amey.pasarkar@princeton.edu | [amepas.github.io](https://amepas.github.io)

## EDUCATION

---

### Princeton University

*PhD candidate in Machine Learning*, Advisor: Adji Bouso Dieng

- Awarded 2022 NSF-GRFP Fellowship

Princeton, NJ  
May 2027 (*expected*)

### Columbia University

*Bachelor of Science in Operations Research*

- Cumulative GPA: 4.02/4.00, Magna Cum Laude
- Finalist for Computing Research Association Undergraduate Research Award

New York, NY  
2018 – 2022

## EXPERIENCE

---

### Princeton Department of Computer Science, Vertaix

*PhD Student*

- Developing scalable dataset curation algorithms for billion-scale training corpora to detect biases, identify memorization, and improve interpretability of generative AI models.
- Designing information-theoretic optimization methods for vision model self-supervised learning.
- Building post-training algorithms for LLMs, including data-efficient instruction tuning and reinforcement learning policy optimization methods to improve LLM reasoning and alignment.

Princeton, NJ  
March 2023 – Present

### Amazon, FireTV Science

*Research Scientist Intern*

- Trained LLMs to improve video search query understanding on the Amazon FireTV.
- Designed and implemented training and evaluation pipelines for LLMs ranging from 0.6 to 70 billion parameters.
- Applied Chain-of-Thought and Direct Preference Optimization to improve model performance, achieving 5× lower inference latency and 10% higher accuracy compared to the baseline production model.

Sunnyvale, CA  
May 2025 – Aug 2025

### Columbia Department of Computer Science, Itsik Pe'er Lab

*Research Assistant*

- Developed probabilistic dynamical models for noisy, high-dimensional time-series and spatial datasets.
- Created new class of latent-variable mixture models to represent structured, spatially dependent distributions.
- Derived scalable approximate inference algorithms for these models by extending techniques like variational inference and expectation-maximization.

New York, NY  
Aug 2019 – Dec 2021

### Amazon, AWS Safety Engineering

*Software Development Engineer Intern*

- Built and deployed pipeline for performing context-based queries on documents detailing incidents in all Amazon services. Model results helped identify vulnerabilities in services and improve their reliability.

Seattle, WA  
May 2021 – Aug 2021

### Columbia Department of Computer Science, Internet Real Time Lab

*Research Assistance, Advisor: Henning Schulzrinne*

- Analyzed FCC data to determine how COVID-19 affected internet usage and network performance in the US.

New York, NY  
Sep 2020 – Dec 2020

## SELECTED PUBLICATIONS

---

- **Amey P. Pasarkar**, Adji Bouso Dieng. “The Vendiscope: An Algorithmic Microscope for Data Collections.” *arXiv* preprint (under review), 2025.
- **Amey P. Pasarkar**, Adji Bouso Dieng. “Cousins of the Vendi Score: A Family of Similarity-Based Diversity Metrics for Science and Machine Learning.” *AISTATS*, 2024.
- **Amey P. Pasarkar**, Gianluca M. Bencomo, Simon Olsson, Adji Bouso Dieng. “Vendi sampling for molecular simulations: Diversity as a force for faster convergence and better exploration.” *Journal of Chemical Physics*, 2023.
- **Amey P. Pasarkar**, Tyler A. Joseph, and Itsik Pe'er. “Directional Gaussian Mixture Models of the gut microbiome elucidate microbial spatial structure.” *mSystems*, 2021.
- Jessica De Oliveira Moreira, **Amey P. Pasarkar**, ..., Henning G. Schulzrinne. “Social Distancing and the Internet: What Can Network Performance Measurements Tell Us?” *The Research Conference on Communications, Information, and Internet Policy* 48, 2021.