Amey Pasarkar

amey.pasarkar@princeton.edu | amepas.github.io

EDUCATION

Princeton University, Quantitative and Computational Biology — **Princeton, NJ**
PhD candidate, Advisor: Adji Bousso Dieng — *2027 (expected)*

- Awarded 2022 NSF-GRFP Fellowship

Columbia University, Fu Foundation School of Engineering — **New York, NY**
Bachelor of Science in Operations Research — *Aug 2018 – May 2022*

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

RESEARCH EXPERIENCE

Princeton Department of Computer Science, Vertaix — **Princeton, NJ**
PhD Student — *March 2023 – Present*

- Research on methods for curating and evaluating datasets for training large-scale AI models. Focus on enabling data-efficient training and improving model generalization abilities. Past projects included benchmark image, text, and protein datasets.
- Designing faster sampling algorithms through flexible regularization schemes, with applications in image generative modeling, molecular dynamics, and material design.

Columbia Department of Systems Biology, Itsik Pe'er Lab — **New York, NY**
Research Assistant — *Aug 2019 – Aug 2021*

- Developed probabilistic methods for analyzing spatial and temporal microbiome data.
- Created state-space models to recover microbial dynamics from noisy and unstructured sequencing data.

Columbia Department of Computer Science, Internet Real Time Lab — **New York, NY**
Research Assistance, Advisor: Henning Schulzrinne — *Sep 2020 – Dec 2020*

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

PROFESSIONAL EXPERIENCE

Amazon, AWS Safety Engineering — **Seattle, WA**
Software Development Engine — *May 2021 – Aug 2021*

- Built and deployed NLP 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.

Clinton Group — **New York, NY**
Activist Hedge Fund Intern — *June 2019 – Aug 2019*

- Utilized Python and Excel to support investment analysis through data-driven analytical insights and visualizations.

PUBLICATIONS & PRESENTATIONS

- **Amey P. Pasarkar**, Adji Bousso 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 Bousso 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.
- **Amey P. Pasarkar**, Tyler A. Joseph, and Itsik Pe'er. "Probabilistic modelling of the gut microbiome's spatial structure using directional Gaussian mixture models." *Probabilistic Modeling in Genomics*, 2021 (Poster presentation).
- 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.
- Tyler A. Joseph, **Amey P. Pasarkar**, and Itsik Pe'er. "Efficient and Accurate Inference of Mixed Microbial Population Trajectories from Longitudinal Count Data." *Cell Systems*, 2020.
- Tyler A. Joseph, **Amey P. Pasarkar**, and Itsik Pe'er. "Efficient and accurate inference of microbial trajectories from longitudinal count data." *RECOMB*, 2020.