

Rajiv Movva

rmovva@berkeley.edu | [Webpage](#) | Updated 02/2026
Interests: AI for science; human-centered AI; healthcare

EDUCATION

University of California, Berkeley Ph.D., Computer Science. Advisor: Emma Pierson.	Jan 2025–present
Cornell University Ph.D., Computer Science. Advisor: Emma Pierson. Note: Transferred to Berkeley in third year.	2022–2024
Massachusetts Institute of Technology B.S., Computer Science. GPA: 4.9/5.0. Minors: Biology; Women’s and Gender Studies.	2018–2022

HONORS

Best Findings Paper, Honorable Mention, ML4H	2023
Digital Life Initiative Doctoral Fellow	2023
NSF Graduate Research Fellow	2022
Best Student Paper, FAccT	2022
Finalist, Hertz Fellowship	2022
Best Paper, BlackboxNLP Workshop @ EMNLP	2020
Finalist, Regeneron Science Talent Search	2018
Platinum, USA Computing Olympiad	2016

INDUSTRY EXPERIENCE

Microsoft Research Research Intern, FATE Montréal. Mentors: Su Lin Blodgett, Noura Farra. Evaluating LLM annotation of responsible AI harms.	2024
Apple Research Intern, Siri. Mentors: Shayne Longpre, Jinhao Lei, Ajay Gupta. Quantization, pruning, & distillation for efficient BERT inference.	2021
NVIDIA Research Intern, Genomics. Mentor: Avantika Lal. GPU acceleration for fast single-cell ATAC-seq analysis.	2020
Genesis Therapeutics Research Intern. Mentors: Ben Sklaroff, Evan Feinberg. Graph neural networks for ligand-protein binding prediction.	2019

ADDITIONAL RESEARCH EXPERIENCE

Carbin Lab, MIT CSAIL Neural network pruning. Mentor: Jonathan Frankle.	2020-22
D'Ignazio Lab, MIT DUSP Participatory NLP for gender activism. Mentor: Harini Suresh.	2021-22
Badran Lab, Broad Institute Directed evolution for genetic code expansion. Mentor: Erika DeBenedictis.	2018-19
Kundaje Lab, Stanford Genetics Deep learning for regulatory genomics. Mentor: Peyton Greenside.	2017-19

PAPERS

Also see [Google Scholar](#). Asterisks* denote equal authorship.

15. **Rajiv Movva**, Smitha Milli, Sewon Min, Emma Pierson. [What's In My Human Feedback? Learning Interpretable Descriptions of Preference Data](#). *ICLR*, 2026. [Oral presentation](#) (top ~1% of submissions).
14. **Rajiv Movva***, Kenny Peng*, Nikhil Garg, Jon Kleinberg, Emma Pierson. [Sparse Autoencoders for Hypothesis Generation](#). *ICML*, 2025.
13. Kenny Peng*, **Rajiv Movva***, Jon Kleinberg, Emma Pierson, Nikhil Garg. [Use Sparse Autoencoders to Discover Unknown Concepts, Not to Act on Known Concepts](#). Preprint, 2025.
12. Severin Engelmann*, **Rajiv Movva***. [LLMs' Pluralistic Compatibility](#). *European Workshop on Algorithmic Fairness*, 2025.
11. **Rajiv Movva**, Pang Wei Koh, Emma Pierson. [Annotation alignment: Comparing LLM and human annotations of conversational safety](#). *EMNLP* 2024.
10. Emma Pierson*, Divya Shanmugam*, **Rajiv Movva***, Jon Kleinberg* et al. [Using Large Language Models to Promote Health Equity](#). *New England Journal of Medicine AI*, 2025.
9. Divya Shanmugam, Monica Agrawal, **Rajiv Movva**, Irene Y. Chen, Marzyeh Ghassemi, Maia Jacobs, Emma Pierson. [Generative AI in Medicine](#). To appear, *Annual Review of Biomedical Data Science*, 2025.
8. **Rajiv Movva***, Sidhika Balachandar*, Kenny Peng*, Gabriel Agostini*, Nikhil Garg, Emma Pierson. [Topics, Authors, and Institutions in Large Language Model Research: Trends from 17K arXiv Papers](#). *NAACL* 2024.
7. **Rajiv Movva***, Divya Shanmugam*, Kaihua Hou, Priya Pathak, John Gutttag, Nikhil Garg, Emma Pierson. [Coarse race data conceals disparities in clinical risk score performance](#). *MLHC* 2023; *ML4H* 2023. [Honorable Mention, Best Findings Paper, ML4H](#).
6. Harini Suresh, **Rajiv Movva**, Amelia Dogan, Rahul Bhargava, Isadora Cruxên, Ángeles Mar-

- tinez Cuba, Giulia Taurino, Wonyoung So, Catherine D'Ignazio. [Towards Intersectional, Feminist, Participatory ML: A Case Study in Supporting Femicide Counterdata Collection](#). *FAccT 2022*. **Best Student Paper**.
5. **Rajiv Movva***, Jinhao Lei*, Shayne Longpre, Ajay Gupta, Chris DuBois. [Combining Compressions for Multiplicative Size Scaling on Natural Language Tasks](#). *COLING 2022*.
 4. **Rajiv Movva**, Jason Zhao. [Dissecting Lottery Ticket Transformers: Structural and Behavioral Study of Sparse Neural Machine Translation](#). *BlackboxNLP @ EMNLP 2020*. **Best Paper**.
 3. **Rajiv Movva**, Jonathan Frankle, Michael Carbin. [Studying the Consistency and Composability of Lottery Ticket Pruning Masks](#). *ICLR Workshop on Science and Engineering of Deep Learning 2021*.
 2. **Rajiv Movva**, Peyton Greenside, Georgi K. Marinov, Surag Nair, Avanti Shrikumar, Anshul Kundaje. [Deciphering regulatory DNA sequences and noncoding genetic variants using neural network models of massively parallel reporter assays](#). *PLoS ONE*, 2019.
 1. Remzi Celebi, Oliver Bear Don't Walk IV, **Rajiv Movva**, Semih Alpsoy, Michel Dumontier. [In-silico Prediction of Synergistic Anti-Cancer Drug Combinations Using Multi-omics Data](#). *Scientific Reports*, 2019.

OTHER WRITING

- **Rajiv Movva.** [GenAI's Burden of Authenticity](#). Digital Life Initiative, 2024.
- **Rajiv Movva**, Pang Wei Koh, Emma Pierson. [Using unlabeled data to enhance fairness of medical AI](#). *Nature Medicine (News & Views)*, 2024.

TALKS & PRESENTATIONS

What's In My Human Feedback?

Meta, Mechanistic Interpretability Reading Group
[UK AI Security Institute](#), Societal Impacts Seminar

February 2026

February 2026

Sparse Autoencoders for Hypothesis Generation

[ML-Economics Summer Conference](#), UChicago Booth
[Transluce](#)

August 2025

August 2025

Sociotechnical Alignment Center, Microsoft Research
Center for Human-Compatible AI, UC Berkeley

July 2025

April 2025

Equitable Machine Learning in Healthcare

International Conference on Health Policy Statistics (ICHPS), San Diego, CA

January 2025

Goals and Challenges for Pluralistic AI

Digital Life Initiative, Cornell Tech

April 2024

Topics, Authors, and Institutions in LLM Research

Text as Data (Poster), UMass Amherst
Science of Science Journal Club, Cornell University
[Data Skeptic Podcast](#)

November 2023

October 2023

September 2023

ADVISING / MENTORSHIP

[Aditya Dawar](#)

Fall 2025–

[Qingpeng Kong](#)

Fall 2025–

[Vatsal Baherwani](#)

Summer 2025–

SERVICE

Reviewer, ICLR

2026

Reviewer, FAccT

2023, 2024, 2025

Reviewer, ACL Rolling Review

2023, 2024, 2025

Reviewer, NeurIPS Behavioral ML Workshop

2024