

# Rajiv Movva

rmovva@berkeley.edu | [Webpage](#) | Updated 12/2025

**Interests:** human-centered AI; AI for science; healthcare

## EDUCATION

---

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

## INDUSTRY EXPERIENCE

---

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

## 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

## 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](#). Preprint, 2025.
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 Guttag, 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 Cruxê, Ángeles Martínez Cuba, Giulia Taurino, Wonyoung So, Catherine D'Ignazio. [Towards Intersectional, Feminist, Participatory ML: A Case Study in Supporting Feminicide 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

<b>Sparse Autoencoders for Hypothesis Generation</b>	
<a href="#">ML-Economics Summer Conference</a> , UChicago Booth	August 2025
<a href="#">Transluce</a> , San Francisco, CA	August 2025
Sociotechnical Alignment Center, Microsoft Research	July 2025
Center for Human-Compatible AI, UC Berkeley	April 2025
<b>Equitable Machine Learning in Healthcare</b>	
International Conference on Health Policy Statistics (ICHPS), San Diego, CA	January 2025
<b>Goals and Challenges for Pluralistic AI</b>	
Digital Life Initiative, Cornell Tech	April 2024
<b>Topics, Authors, and Institutions in LLM Research</b>	
Text as Data (Poster), UMass Amherst	November 2023
Science of Science Journal Club, Cornell University	October 2023
<a href="#">Data Skeptic Podcast</a>	September 2023

## ADVISING / MENTORSHIP

Aditya Dawar Fall 2025–  
Qingpeng Kong Fall 2025–  
Vatsal Baherwani Summer 2025–

## SERVICE

Reviewer, ICLR	2025
Reviewer, FAccT	2023, 2024, 2025
Reviewer, ACL Rolling Review	2023, 2024, 2025
Reviewer, NeurIPS Behavioral ML Workshop	2024