

Gautam Machiraju

6044 College Ave, Apt C | Oakland, CA 94618 | 614-599-6288

gmachi@stanford.edu ♦ [gmachiraju.github.io](https://github.com/gmachiraju) ♦ [🐦](#) [📘](#) @gmachiraju

Education

Stanford University

2018 – 2024

Ph.D. Biomedical Informatics — Department of Biomedical Data Science (BDS)

Advisors: Parag Mallick (Radiology), Christopher Ré (CS)

Committee: Sylvia Plevritis (BDS, Radiology), James Zou (BDS)

Thesis title: Deep learning for de novo discovery of verifiable pathologies

Research agenda: Building AI-driven data copilots for scientific discovery

University of California, Berkeley

2012 – 2016

B.A. Applied Mathematics, Minor in Bioengineering

Interests: Mathematical Biology (elected degree concentration) and numerical analysis with particular emphasis in differential equations, optimization, & algorithms

Recent Work Experience

Graduate Student Researcher @ Mallick & Ré Labs | Stanford AI Lab (SAIL)

2018 – Present

- Equipping feature attribution capabilities to foundation models (LLMs, VLMs, etc.) that operate on coarsely labeled high-dimensional data — e.g. localizing tumors in gigapixel histopathology, identifying binding sites in protein structures, & retrieving salient text in documents [IMLH'23, ICML'24]
- Evaluation frameworks for feature attribution capabilities of partial-context models [ECCV'22]
- Currently deploying methods for AI-driven knowledge discovery: identifying novel prognostic biomarkers of cancer progression in spatial biology
- Supported projects on: LLM-driven text mining of PubMed literature for clinical biomarkers, equipping feature attribution capabilities to DNA language models
- *Keywords:* AI4science, explainability & interpretability, foundation models, geometric ML, long-range modeling, weak supervision

Graduate Research Consultant @ Thieme & Geldsetzer Labs | Stanford Medicine

2021 – present

- Moonlighting projects to build out AI for monitoring global health:
 1. App-based mpox detection with mobile-friendly vision models trained on social media corpus of skin lesion images [NatMed'22]
 2. Deep regression models to predict maternal & child health outcomes of remote villages in low- & middle-income countries using satellite imagery, government surveys, & other remote sensing data
 3. Clinical-grade VLMs built to pass USMLE medical board exams, supporting long text prompts
- *Keywords:* edge computing, foundation models, multimodality

Student Researcher @ IBM Research

Summer – Fall 2023

- Stanford-based collaboration to enable visually grounded foundation models for downstream use in clinical settings (AR, ambient intelligence); open vocabulary object detection to enable memory surrogates; fact-checking VLMs for improved image tagging & retrieval
- *Keywords:* foundation models, mulitmodality, trustworthiness

Graduate Teaching Assistant @ Department of BDS, Stanford University

2019 – 2020

- BIOMEDIN 214 (Russ Altman): Representations & Algorithms for Computational Molecular Biology
- Fall 2019: Developed coding assignments & exams, led weekly office hours, delivered lectures on CS fundamentals, Python, & deep learning; received **teaching award** from student evals
- Fall 2020: Transitioned curriculum to remote-friendly format during COVID-19 pandemic

Bioinformatics Research Assistant @ Mallick Lab, Stanford Radiology

2016 – 2018

- Physics-based models (ODEs) for biomarker shedding kinetics using tumor microenvironment data; enabled simulations to assess serum biomarker viability [NatSciRep'20]

- Supported projects on: anomaly detection for multivariate time-series (multi-omics + wearables) to identify drivers of viral disease onset, NLP of PubMed articles to construct biomarker databases
- *Keywords*: physics-based models, numerical methods, NLP, time-series analysis




Bioengineering SDE Intern @ Strateos (formerly Transcriptic)

Summer 2016


Software engineering & scientific computing to refine robotic platform for wet-lab automation; projects in convex optimization of robotic gantry movement and queueing & search of platform's job requests

Current Projects

Thesis work:

- Gautam Machiraju, Christopher Ré, Parag Mallick. Explainable Machine Learning for Digital Pathology & Spatial Biology. In preparation for *Nature*.
 Reviewing state of the art methods to classify and explain patient diagnoses and prognoses
- Gautam Machiraju, Neel Guha. Toward AI-driven Data Copilots for Scientific Inquiry. In preparation for the *Hazy Research Blog* (<https://hazyresearch.stanford.edu/blog>).
 Blog post for the Hazy Research Blog, discussing prospector heads (ICML'24) and their applications to scientific benchmarks, as well as knowledge discovery in spatial biology and human/agent game-play characterization in chess
- Gautam Machiraju, Amir-Hossein Karimi, James Zou, Sylvia Plevritis, Fiona Ginty, Christopher Ré, Parag Mallick. AI-driven Biomarker Discovery for Spatial Biology. In preparation for *Nature Methods*
 Applying prospector heads (ICML'24) to spatial biology in the search for clinically relevant biomarkers of cancer progression

Non-thesis work:

- Haojie Wang, Gautam Machiraju, Ali Lenjani, Turan Orenbas, Luke Carani, Pascal Geldsetzer. Predicting Maternal and Child Health Indicators using Publicly Available Remote Sensing, Survey, and Geo-tagged Data. *In review*.
 An applied vision project to perform global health monitoring; aim to predict maternal and child health outcomes of Earth's remote villages through the use of Landsat satellite imagery (Google Earth Engine), government surveys (Demographic and Health Surveys Program), and other geo-tagging and remote sensing data

Peer-reviewed Works

Note: additional writing artifacts & pre-prints can be found on personal website.

Feature Attribution for High-dimensional Scientific Data (thesis work):

1. Gautam Machiraju, Christopher Ré, Parag Mallick. Prospector Heads: Generalized Feature Attribution for Large Models & Data. *Compound AI Systems Workshop*, 2024.
2. Gautam Machiraju, Alexander Derry, Arjun Desai, Neel Guha, Amir-Hossein Karimi, James Zou, Russ Altman, Christopher Ré, Parag Mallick. Prospector Heads: Generalized Feature Attribution for Large Models & Data. *International Conference on Machine Learning (ICML)*, 2024.
3. Gautam Machiraju, Arjun Desai, James Zou, Christopher Ré, Parag Mallick. Prospectors: Leveraging Short Contexts to Mine Salient Objects in High-dimensional Imagery. *International Conference on Machine Learning (ICML) workshop on Interpretable Machine Learning for Healthcare (IMLH)*, 2023.
4. Gautam Machiraju, Sylvia Plevritis, Parag Mallick. A Dataset Generation Framework for Evaluating Megapixel Image Classifiers & their Explanations. *European Conference on Computer Vision (ECCV)*, 2022.

Training & Fine-tuning Foundation Models for Biomedicine (non-thesis work):

5. Alexander Henry Thieme, Tahir Miriyev, Gautam Machiraju, et al. A Medical Vision-Language Model Trained on a Large Dataset of Open Access Medical Images. *In Review*.
6. Varun Tandon, Gautam Machiraju, Parag Mallick. Grammar Matters: Exploring Grammatical Variation's Role in Improving Fine-Tuned LLMs for Biomedical Relation Extraction. *In Review*.

ML for Public Health & Clinical Decision Support (non-thesis work):

7. Alexander Henry Thieme, Yuanning Zheng, Gautam Machiraju, et al. Development and Evaluation of an Image-based Deep Learning Algorithm to Classify Skin Lesions from Mpox Virus Infection. *Nature Medicine*, 2023.
8. Minh Nguyen, Conor Corbin, Tiffany Eulalio, Nicolai Ostberg, Gautam Machiraju, Ben Marafino, Michael Baiocchi, Christian Rose, Jonathan Chen. Developing Machine Learning Models to Personalize Care Levels among Emergency Room Patients for Hospital Admission. *Journal of the American Medical Informatics Association (JAMIA)*, 2021.

Mathematical & Statistical Modeling for Cancer Systems Biology (non-thesis work):

9. Hunter Boyce, Gautam Machiraju, Parag Mallick. Spatial Statistics for Spatial Biology. *In Review*.
10. Gautam Machiraju, Parag Mallick, Hermann Frieboes. Multicompartment Modeling of Protein Shedding Kinetics During Vascularized Tumor Growth. *Nature Scientific Reports*, 2020.

Awards, Scholarships, & Fellowships

- Creativity in Research Scholars (CIRS) Program, Stanford d.school (2023 cohort)
- Canary-ACED Graduate Fellowship (100% support) via International Alliance for Cancer Early Detection (2022-2023)
- Stanford Data Science Scholarship (50% support) via Stanford Data Science Institute (2021-2022, 2023-2024)
- 2022 HAI Google Cloud Credits Award (\$40K) for “Algorithmically identifying histo-molecular biomarkers of cancer progression” (2022-2023)
- National Library of Medicine (NLM) Training Grant (100% support) via the NIH (2018-2022)

Academic Teaching, Service, & Leadership Experience

- **Research Mentor**, SMASH Rising. Co-mentoring four undergraduate students of underrepresented backgrounds in STEM. Taught programming and proteomics background material; led research to identify proteins shed into blood with in-house mouse xenografts. June 2022 – August 2022.
- **Reviewer**, European Conference on Computer Vision (ECCV), 2022
- **Reviewer**, Annual Biomedical Research Conference for Minority Students (ABRCMS), 2022
- **Mentor**, Stanford ADVANCE Mentor Collective Program (in partnership with Mentor Collective). Advising mentees from around the US who are applying to colleges or graduate schools. Part of the Stanford ADVANCE volunteer group. Received training and certification to be on-call as a mentor when requested. June 2021 – Present.
- **Co-Czar**, Biomedical Informatics Training Program. Elected student leader and liaison working on recruitment, admissions, student resources and events (e.g. departmental Roundtables with invited speakers), curriculum, DEI efforts (e.g. Antiracism Book Clubs), socials, etc. Host weekly office hours for student feedback. Sit on departmental executive committee. Estimated 500 hours of service to the department. October 2020 – October 2021.

- **Research Mentor**, SMASH Rising. Co-mentoring four undergraduate students of underrepresented backgrounds in STEM. Students are exposed to and work toward identifying candidate protein biomarkers of cancer through bioinformatics approaches. Gave daily lectures on systems biology, multi-omics, mass spectrometry, data wrangling, and data visualization. Introduce Python via Google Colab notebooks. Helped form research questions culminating in research project. June 2021 – August 2021.
- **Panelist**, Stanford ADVANCE - Rotations Workshop. July 9, 2021.
- **Committee Member**, DBDS Faculty JEDI Committee. See details below for efforts related to justice, equity, diversity, and inclusion (JEDI). September 2020 – September 2021.
- **Committee Member**, Student DEI Committee of the Biomedical Informatics Training Program. Student and faculty efforts to diversify admissions and culture at Stanford's Department Biomedical Data Science (DBDS). Spearheaded bi-annual departmental DEI Town Halls, antiracism book clubs, the funding of Women and Non-binary Students of Biomedical Data Science, an application review program for prospective students, and a *BMI Bootcamp* to onboard a diverse incoming MS and PhD student body through informatics and career development workshops. June 2020 – September 2021.
- **Mentor**, BIODS 360: Inclusive Mentorship in Data Science. Working with undergraduates at other institutions to introduce applied data science concepts and career paths in industry and academia. Tailoring weekly exercises to learn concepts. March 2021 – June 2021.
- **Mentor**, ADVANCE Undergraduate Institute (AUI). Mentored future Stanford applicants in interview preparation and application reviews. April 2021.
- **Graduate Teaching Assistant**, BIOMEDIN 214 (Representations & Algorithms for Computational Molecular Biology). Worked with Prof. Russ Altman to update syllabus, headed office hours and review sessions, developed exam questions and code for projects. Designed and gave two lectures: Python for bioinformatics and deep learning methods applied to biological data. Received teaching award from the Stanford Center for Professional Development (SCPD). Autumn 2019 & 2020.
- **Reviewer**, Deep Learning for Genomics, IEEE/ACM Transactions on Computational Biology and Bioinformatics 2020
- **Reviewer**, ML4Health workshop, NeurIPS 2019
- **Trainee Panelist**, AI in Radiology Research, Stanford Radiology Joint Research Retreat; Oct 29, 2019
- **Student Planning Committee Member**, Stanford Department of Biomedical Data Science (BDS) Scientific Retreat 2019; Sept 27, 2019
- **Volunteer**, The Early Detection of Cancer Conference, 2019; Sept 24-25, 2019.
- **Alumni Panelist**, Mathematics Undergraduate Student Association (MUSA), UC Berkeley; May 3, 2018
- **Organizing Committee Member**, Biomedical Computation at Stanford (BCATS) Symposium 2018
- **Research Mentor**, Canary Cancer Research Education Summer Training (CREST) Program, 2018. Mentoring students to aid in lab's NLP project (see below), ultimately with a submission to the Siemens Competition.
- **Organizing Committee Member**, Stanford Radiology Joint Research Retreat 2017
- **Research Mentor**, Canary Cancer Research Education Summer Training (CREST) Program, 2017. Culminated in building Natural Language Processing pipeline for PubMed data extraction, with ultimate goal of creating a public database of disease biomarkers.

Entrepreneurship Experience & Incubator Membership

-
- **Sutardja Center for Entrepreneurship and Technology, UC Berkeley**, January – May 2017
 - **Free Ventures, UC Berkeley**, January – May 2015
 - **Catalyst@Berkeley**, September – December 2014

Relevant Skills

Programming	Python (Pytorch, Tensorflow), shell [extensive] R, MATLAB [intermediate] SQL, CUDA, C++, Java, Spark, Scala, Javascript, HTML/CSS [basic]
Workflow	Cluster (Slurm), Cloud (GCP, Kubernetes), integrations (VSCode, W&B, Jupyter, rmate)
Design	L ^A T _E X, vector graphics (BioRender, Adobe suite)
Mentorship	Mentored 15 URM trainees on independent research projects

Further Inquiry

❗ For more details regarding research projects, publication & preprints, industrial work experience, academic service, coursework (in BDS, CS, EE, Stats, Maths, etc.), and entrepreneurial experience, please visit my Linkedin page (<https://www.linkedin.com/in/gmachiraju>) or personal website (gmachiraju.github.io).