Pritika Vig

Cambridge, MA • pritikav@mit.edu • 603-724-5904

Research Interests

Studying the causal and geometric organization of representations in self-supervised models to understand how inductive biases shape interpretable features that generalize to rare and data-scarce biomedical domains.

Education

Massachusetts Institute of Technology — *MBA, Sloan School of Management* *2024 – 2026 (Expected)*

- Cross-registered in EECS coursework: Deep Learning (6.7960), Machine Learning for Healthcare (6.7930/HST.956), Modeling with Machine Learning (6.C51/6.C511)
- Healthcare Certificate: Principles and Practices of Drug Development (15.136J/HST.920J/IDS.620J), Healthcare Lab (15.777)

Dartmouth College — *BA in Computer Science* *2013 – 2017*

- Thesis on efficient motion planning algorithms for rigid bodies in 2D environments with polygonal obstacles.

Research Experience

Dana-Farber Cancer Institute / Harvard Medical School *June 2025 – Present*
Research Assistant
Advisor: Dr. William Lotter, Department of Data Science

- Leading first-author project investigating if vision transformers trained on static images implicitly learn the underlying temporal dynamics of disease; targeting ICML 2026 submission.
- Developed a manifold analysis framework benchmarking 6 foundation models (e.g., UNI2, CONCH, GigaPath) using diffusion pseudotime; validated findings via label-shuffle nulls and stage-permutation analysis to prove learned geometry encodes true biological trajectory.
- Discovered that disease progression information peaks at intermediate network depths (60-75% through layers) before plateauing. Established that Kendall's tau on reference progressions is a strong predictor of performance on held-out staging tasks (Pearson r=0.705, $p < 0.001$), demonstrating that geometric quality metrics can forecast model generalization.

MIT Institute for Medical Engineering & Science (IMES) *May 2025 – Present*
Research Assistant
Advisor: Dr. Leo Anthony Celi, Computational Physiology Lab

- Designing engineering framework treating curiosity and humility as computational primitives for clinical decision support, enabling systems to learn when to seek information versus defer decisions
- Proposed benchmarking against OpenAI's HealthBench, achieving ∼40% improvement on challenging diagnostic cases through principled uncertainty quantification
- Coauthor on project on sycophancy bias in clinical LLMs: developing synthetic patient profile framework with cookie-based testing to quantify if models alter medical recommendations based on patient demographics rather than clinical evidence

Publications and Manuscripts

Under Review

- Abulibdeh R, Agyemang GO, Ameen S, Cadillac L, Celi LA, Maciejewski M, Samsel K, **Vig P**. Moving beyond can we to should we: an ethics by design framework for healthcare AI. Humanit Soc Sci Commun. Under review.
  Equal contribution among authors.
- Arslan J, Benke K, Cajas Ordoñez SA, Castro R, Celi LA, Cruz-Suarez GA, Delos Reyes R, Engelmann J, Ercole A, Hilel A, Kalla M, Kinyera L, Lange M, Lunde TM, Meni MJ, Ocampo Osorio F, Premo AE, Sedlakova J, **Vig P**. An engineering framework for curiosity-driven and humble AI in clinical decision

support. BMJ Health Care Inform. Under review.
Equal contribution among authors.

In Preparation

- **Vig P**, Lotter W. Learning temporal order without supervision: emergent progression in pathology foundation models. Manuscript in preparation for ICML 2026.

Professional Experience

Google (YouTube) — *Staff Software Engineer (promoted from SE II, SE III, Senior) Technical Lead, YouTube Music Identity* *Sept 2017 – Aug 2024*

- **Distributed Systems & Scalability:** Architected identity and personalization infrastructure supporting scale from 7M to 100M users; designed low-latency data retrieval systems for user profiles adopted across YouTube Kids and Gaming.
- **Data Engineering on Large Datasets:** Led the backend engineering for "Music Recap," building data pipelines to aggregate and analyze year-long longitudinal user logs for 100M+ users to generate personalized, dynamic, shareable insights.
- **Algorithmic Optimization:** Diagnosed a content availability failure affecting 1.2M tracks/day; designed and deployed a fallback logic algorithm via controlled A/B testing that reduced session errors by 40%, demonstrating rigorous experimental validation in production.
- **Privacy Engineering & Governance:** Collaborated with legal counsel to operationalize "Privacy by Design" principles for managing user consent and history; built music history infrastructure compliant with GDPR regulations.
- **Technical Leadership:** Guided architectural decisions through "Area Tech Lead" reviews (YouTube's highest technical council); mentored a team of 4 engineers on system design, code quality, and reliable release practices for high-availability services. Presented two technical talks to YouTube teams, sharing design architectural system design choices.

Google Core Infrastructure Privacy — *Senior Privacy Engineer* *Jan 2021 - June 2021 (Rotation)*

- Trained on GDPR and Privacy by Design principles; reviewed and approved infrastructure launches for compliance
- Performed privacy audits and produced data access recommendations adopted by YouTube Search, Gmail, and Google Search

Teaching Experience

Teaching Assistant — Artificial Intelligence (CS 76), Dartmouth College (2016-2017): Led weekly office hours, graded assignments
Teaching Assistant — Software Principles and Design (CS 50), Dartmouth College (2016-2017): Led weekly recitations for 7 students, graded assignments, held office hours
Technical Team Lead, Google/YouTube Music (2020-2024): Mentored 4-6 engineers through design reviews and career development; delivered technical talks on privacy-compliant ML architectures
Software Engineering Intern Mentor, Google (2018-2020): Coached two interns through full conversion process including project design, implementation oversight, and interview preparation

Service, Mentorship & Leadership

MIT Women in EECS (2024): Mentor and gave talk to the undergrad women in CS club
Vamonos Outside (2022-2024): Mountain biking coach for Latino high schoolers to increase outdoor access
Oregon Adaptive Sports (2019-2022): Volunteer instructor for adaptive skiing programs for athletes with a variety of disabilities (blindness, cognitive, paralysis, etc)
Women@YouTube (2019-2024): Mentor for rising female engineers at YouTube, represented Google at Grace Hopper in outreach

Awards and Recognition

MIT Sloan Dean's Fellow (2024): Merit based scholarship
Scott Silver Award (2019): YouTube's recognition for exceptional technical contributions directly from VP of Engineering
YouTube Northstar Awards (2018, 2020): For embodying the values of YouTube
Launch Honors: Coachella Livestream, Google Play Library Migration, YouTube Recap

Technical Skills

Machine Learning: PyTorch, TensorFlow, Vision Transformers (CONCH, UNI, DINOv2), Self-supervised Learning, Manifold Learning
Medical AI: Whole Slide Image Analysis, Disease Progression Modeling, Survival Analysis, Histopathology Pipelines
Languages: C++, Python, Java, R
Infrastructure: Offline queue and batch processing, database management, Conda, Docker, GCP
Technical Leadership: Code review, project tracking, identifying blockers, design document generation and review, blameless postmortems