Vijay Vankadaru

(510) 710-6220 | vankadaruvijay@berkeley.edu | linkedin.com/in/vijay-vankadaru | vijay-vankadaru.github.io | Google Scholar

Education

University of California, Berkeley **Expected May 2027**
Master of Data Science, Machine Learning emphasis, GPA: 4.0 *Berkeley, CA*
- Coursework: NLP with Deep Learning, Computer Vision, Applied Machine Learning, Causal Inference, Data Engineering, Statistics

University of California, Riverside **June 2025**
Bachelor of Science in Computer Science *Riverside, CA*

Publications

[1] **V. Vankadaru**, V. DuTell. Vision-Language Models Trust What They Can Discern. *ECCV 2026 (eXCV)*.

[2] **V. Vankadaru**, A. Matthews, T. Roosta, P. Passban. Readable but Not Controllable: Neuron-Level Evidence for Medical LLM Hallucination. *COLM 2026*.

[3] A. Matthews, **V. Vankadaru**, T. Roosta, P. Passban. Rethinking Medical LLM Hallucinations: A System-Level Survey. *Artificial Intelligence in Medicine*, in press.

[4] **V. Vankadaru**, V. DuTell. Believing Is Seeing: How LLaVA Models Weigh a Contradicting Image. Under review, ICLR 2027.

[5] E. S. Danthinne, **V. Vankadaru**, et al. Potential for Age-Based Bias in AI-Informed Neurosurgical Decision-Making: A Sociodemographic Subgroup Assessment Approach for Open-Source Glioma Molecular Subtype Classifiers, in review.

[6] A. Ge, **V. Vankadaru**, X. Liu, et al. Benchmarking Somatic Mutation Representations for Cancer Typing and HRD Prediction, in review.

[7] **V. Vankadaru**. A Recipe for Multimodal Multi-Instance Depression Detection: Frozen Teachers, Schedule-Aware Training, and Audio Contribution Diagnostics. Under review, ACL Rolling Review.

Research Experience

UC Berkeley Graduate Researcher
Advisor: Dr. Vasha DuTell *Apr 2026 – Present*
- Studied how vision-language models (CLIP-LLaVA) weigh an image against contradicting text, applying activation patching and logit lens (TransformerLens, Hugging Face) to measure the decision; models trust whichever input they can discern; first author, ECCV 2026
- Extended the study to three LLaVA models and four kinds of image-text conflict, adding an evaluation on real photographs (NaturalBench) and multiple-testing correction across all reported effects; full paper under review at ICLR 2027

Advisor: Prof. Tanya Roosta *Jan 2026 – Present*
- Led a neuron-level hallucination study of medical LLMs (BioMistral, Meditron, Llama 3.1): linear probes on hidden activations detect hallucination at AUROC 0.77 to 0.86, yet steering those same neurons cannot reliably suppress it; first author, COLM 2026

UCSF M2AI Lab Graduate Researcher **Jun 2026 – Present**
Advisor: Dr. Madhumita Sushil
- Externally validated three open-source MRI classifiers (3D CNN, ViT with ClinicalBERT, CNN with radiomics) on a 3,380-patient cohort and showed that errors AUROC hides are driven by age: specificity 54% for patients under 40 vs 80% for 55 and older
- Scaled ground-truth labels from 442 to 3,380 patients with an LLM extraction pipeline over 9,377 pathology reports (majority vote over three runs, F1 0.96 to 0.97), which powered subgroup tests with bootstrap CIs, permutation tests, and FDR correction

UMSOM Tensor Lab Graduate Researcher **Jun 2026 – Present**
Advisor: Dr. Bradley A. Maron
- Designed Bio MAF v4, a 948-feature interpretable representation of a tumor's somatic mutations, and benchmarked it against mutational signatures and a from-scratch MuAt transformer on 8,800 TCGA tumors across 20 cancer types
- Showed that feature tables with XGBoost beat the transformer (0.738 vs 0.528 balanced accuracy) and transfer to 913 held-out tumors

Industry Experience

Machine Learning Engineer **Jun 2025 – Present**
DASION (clinical AI startup, NSF SBIR Phase I and II) *Claremont, CA*
- Built a multimodal depression detection model on clinical interview audio and transcripts in PyTorch (wav2vec2 and RoBERTa encoders, gated-attention multi-instance pooling), reaching F1 0.87 on DAIC-WOZ with occlusion diagnostics for per-modality attribution [7]
- Fine-tuned BERT for clinical documentation classification and deployed models as AWS inference endpoints serving 5+ institutions

Machine Learning Engineering Intern **Sep 2021 – Jun 2025**
DASION *Claremont, CA*
- Trained ResNet50 transfer-learning classifiers in PyTorch for lung and cardiac imaging (93% diagnostic accuracy, k-fold CV, SHAP)
- Built a bidirectional LSTM with attention over MFCC and prosodic features for voice-based depression screening (85% accuracy)

Service and Skills

- Reviewer: InterpScience (NeurIPS 2026), Actionable Interpretability (COLM 2026), eXCV (ECCV 2026)
- Presenter, DAIH at COLM 2026 | Lead fellow presenter, The Tensor Lab | BlueDot Impact Technical AI Safety course (2026)
- Skills: Python, PyTorch, TransformerLens, Hugging Face, scikit-learn, XGBoost, SQL, C++, R, AWS, Docker, Linux, LaTeX