

Garv Goswami

Mobile : +1-862-220-1040

Personal Website: garvgoswami.tech

Email : garvgoswami@berkeley.edu

Github: <https://github.com/garvcodes>

EDUCATION

- **University of California, Berkeley** B.A. in Computer Science
Expected Graduation: May 2026 *Aug. 2022 – Present*

INDUSTRY EXPERIENCE

- **Phare Bio** Boston, MA
Machine Learning Intern - Skills: Python, PyTorch, Small Molecule Design, Platform Development *September 2025 – Present*
 - **Benchmarking Platform for Molecular Generation:** Designing and deploying a scalable benchmarking platform on an HPC environment to evaluate molecular generative models—including JT-VAE, Transformer-based, and diffusion models—across large-scale runs (100k+ generated molecules)
 - **Computational Analoging for Antibiotic Discovery:** Conducted computational analoging efforts for small-molecule antibiotic programs using **REINVENT4** (generative design / optimization) and **Chemprop** (property prediction), helping prioritize candidate analogs based on predicted activity and developability-relevant properties.
- **Google** Mountain View, CA
Software Engineering Intern - Skills: C++, Experimentation, System Design *May 2025*
 - **Gemini Voice API:** Contributed to a 20% project exploring conversational recall for dementia patients using the Gemini Voice API. Implemented dialogue flows and evaluated engagement to assess the feasibility of AI-driven patient support.
 - **Ad Format Design and Implementation:** Collaborated w/ YT Shorts Ads team to design & analyze results of new content & component placement for Shorts App Ads. Front-end implementation in EML, server-side changes in C++. Utilized Google's internal experimentation platform to conduct gradual rollouts and A/B tests, optimizing ad performance and user engagement.
- **Amazon** Sunnyvale, CA
Software Development Engineering Intern - Skills: Java, System Design, AWS, Testing *May 2024*
 - **Integration & Load Testing:** Created Integration and Load Testing CI/CD frameworks for new FireTV Ad Stack plugins. Used Java to write tests on temporary Lambda compute to validate packages being added to version set.
 - **Validation Pipeline for FireTV Ads Plugins:** Worked on validation pipeline to add packages to FireTV's Setu, Amazon's internal tool for ad customization. Programmed system monitoring tool in Java to obtain CPU, memory, and latency metrics during plugin integration testing across FireTV ad stack, only allowing addition of plugin to library if all benchmarks are met.

RESEARCH EXPERIENCE

- **Innovative Genomics Institute** Berkeley, CA
Student Researcher: Ronda Lab - Skills: Protein Prediction Models, AlphaFold, Wet Lab Testing *Oct 2024 – Present*
 - **Generative Models for De Novo Protein Design:** Fine-tuned ESM-2 on CRISPR-family proteins to generate novel, compact variants of the SpyCas-9 protein using a custom autoencoder model. Integrated AlphaFold for structural validation.
- **University of California, Berkeley / Icahn School of Medicine at Mount Sinai** Remote Collaboration
First-Author: Clinical Trial Enrollment Prediction Preprint *Aug 2024 – Jan 2025*
 - **Clinical Trial Enrollment Prediction (Preprint):** First-authored draft w/ Mount Sinai medical student on EnrollMate, a machine learning framework predicting clinical trial enrollment success.[Draft Link]

RESEARCH PROJECTS

- **Contrastive Alignment of Expression and Copy Number Reveals Dosage-Insensitive Genes in Cancer**
Accepted to ICLR 2026 MLGenX Workshop | PyTorch, Scanpy, infercnvpy
 - **Contrastive Learning for CNV-Expression Concordance:** Built a contrastive learning framework aligning scRNA-seq expression and inferred CNV profiles in a shared latent space to identify dosage-insensitive genes in cancer, achieving **87%+** bidirectional top-1 retrieval accuracy on held-out cells.
- **Fine-Tuning Genomic Language Models for Variant Pathogenicity Prediction**
Under Review at ISMB 2026 | PyTorch, LoRA, Transformers, Mamba
 - **Genomic Language Model Fine-Tuning:** Benchmarked Nucleotide Transformer and Caduceus on ClinVar missense variant classification, showing that variant-position embeddings outperform mean pooling and that LoRA reaches **0.886** validation AUC with only **0.76%** trainable parameters.