

# Hank Sha

San Jose, CA | (510) 320-1116 | hankssha@gmail.com | linkedin.com/in/hank-sha | hanksha.com | Github

## EDUCATION

### University of California, Santa Barbara

*B.S. Statistics and Data Science*

Goleta, CA

Sep 2022 – June 2026

## EXPERIENCE

### Bouken Capital

*Venture Scout*

Remote

Sep 2025 – Apr 2026

- Translated technical diligence into investment recommendations for 20+ early-stage startups, running market sizing, competitive analysis, and 3 founder interviews to brief investment partners.

### UC Riverside

*Machine Learning Research Assistant*

Remote

June 2025 – Mar 2026

- Isolated which traversal strategies hold up for LLM graph reasoning across 12+ controlled experiments.
- Scaled the evaluation pipeline to 10K+ topological graphs, converting a single-graph prototype into a benchmark suite that ran unattended across the full dataset.
- Cut runtime 75% and GPU memory 20% via batched inference, quantization, and CUDA memory tuning.

### ABVC BioPharma

*Clinical Data Scientist Intern*

Remote

Aug 2025 – Nov 2025

- Ramped into clinical trial methodology in 2 weeks by pairing directly with biologists, translating FDA protocol constraints and endpoint definitions into production modeling workflows for Phase II analysis.
- Delivered a patient-dropout risk model at 83% recall (10-fold CV) by engineering 15+ clinical features and applying SMOTE to an imbalanced cohort.
- Recommended against deploying the model after cross-fold variance exposed an underpowered dataset, surfacing a data-leakage defect in my own pipeline and giving the trial team a defensible basis to redirect resources.

### GitRoll

*AI/ML Engineer Intern*

San Francisco, CA

Sep 2024 – Apr 2025

- Shipped a retrieval-augmented LLM pipeline into production analyzing GitHub repositories, generating developer metrics serving 45K+ users.
- Drove classifier precision to 92% by hand-labeling 216+ PR issues and building a human-in-the-loop evaluation framework, then shipped the resulting NLP impact-scoring system into the core production metric.
- Cut LLM token spend 13% by benchmarking model families on cost-versus-latency, then implementing prompt compression and batched inference against the winner.

## PROJECTS

### Project Nexus — Civic engagement track winner, UCSB Data4Good

2026

Python · LangChain · Pinecone · Neo4j

- Converted 500K+ emails into a knowledge graph via LLM entity extraction, chunking strategy, and relationship modeling.
- Shipped a RAG interface that let non-technical judges trace hidden communities and generate evidence-backed insights against live data under hackathon time constraints.

### Xeno CLI — open-source auto coding workflow

2026 – present

Python · LangGraph · Docker · Tree-sitter

- Orchestrated 6 specialized agents through a LangGraph state machine with Docker-sandboxed execution, enabling autonomous multi-step coding tasks in fully local environments.
- Engineered hybrid local/API routing and prompt caching to secure system prompts and codebase maps from third-party APIs.

### Biotech Intelligence Engine — data platform for biotech-catalyst trading

2026 – present

Python · FastAPI · Celery · TypeScript · Docker

- Ingested 137K+ entries from 5 heterogeneous sources into a unified data feed, deployed via GitHub Actions CI/CD.
- Engineered an in-browser SLM with MedRAG retrieval over drug mechanisms and trial outcomes, integrating Supabase and Upstash for auth and persistent state.

## TECHNICAL SKILLS

**Languages:** Python, SQL, TypeScript, R, bash

**Systems:** Docker, AWS, GCP, CI/CD (GitHub Actions), FastAPI, REST APIs, Redis, Vercel

**Data:** Postgres, Supabase, MongoDB, Pinecone, Neo4j, Pandas, NumPy

**AI/ML:** RAG, LangChain, LangGraph, agent orchestration, evals, prompt engineering, vector databases, PyTorch, Hugging Face, XGBoost, scikit-learn