

ANISH KAMATAM

📍 San Francisco Bay Area 📞 (925) 549-0603 ✉ anishkamatam22@gmail.com 🌐 [linkedin.com/in/anishkamatam](https://www.linkedin.com/in/anishkamatam)

Education

Chico State University

Expected May 2027

B.S. in Computer Information Systems

Chico, CA

- **Coursework:** Programming and Algorithms I & II, Data Structures and Algorithms, Information Systems Design, Linear Algebra

Skills

Python, Java, C++, Go, Rust, JavaScript, TypeScript, SQL (PostgreSQL, MySQL), HTML/CSS, AWS (S3, Lambda, EC2, Amplify, SQS), GCP, Docker, Kubernetes, Git, Node.js, Express, React, Next.js, Flask, Tailwind CSS, Pandas, Pytorch, Scikit-Learn, Keras

Experience

University of California, Berkeley – EECS

Oct 2025 – Present

Undergraduate Research Assistant, Cloud Robotics Lab

Berkeley, CA

- Researching bandwidth-aware cloud robotics using ROS2, and PyTorch for offloading between Jetson Orin Nano and cloud GPUs.
- Deployed transformer-based vision and SLAM models (DINOv2, YOLOv8) with TensorRT for low-latency perception over 5G.
- Built Kubernetes-managed inference pipelines with adaptive compression, model switching, and telemetry profiling for drones.
- Working under Prof. Sylvia Ratnasamy and Dr. Alex Kretsel on SLO-driven scheduling for scalable cloud autonomy.

FoodHealth Company

Nov 2025 – Present

AI/ML Intern

San Francisco, CA

- Trained XGBoost classifiers and regressors on large-scale tabular datasets, optimizing hyperparameters with Optuna for low-latency inference.
- Implemented a feature pipeline with quantile scaling, target-encoded categorical variables, and polynomial interaction terms optimized for tree-based models.
- Built a model audit suite combining SHAP value decomposition, 5-fold cross-validation, and KS-statistic drift checks to validate boosted model stability.

FundMore AI

May 2025 – Nov 2025

AI Engineer Intern

Toronto, Canada

- Built a RAG AI assistant on Google Vertex AI using Sentence-BERT embeddings, Vector Search, and PostgreSQL for analytics.
- Developed multi-tenant inference stack in TypeScript, Angular, and Node.js with Sequelize ORM and Redis caching for queries.
- Implemented function calling and dynamic SQL generation for natural language access to real-time mortgage data.
- Fine-tuned Gemini and PaLM models on financial datasets using Vertex AI Training and reinforcement-style feedback.
- Designed agentic pipelines with Python orchestration and LLM subroutines for retrieval, summarization, and anomaly detection.

AGNI X

Apr 2025 – Sep 2025

Founder & Lead ML Engineer

San Francisco, CA

- Built an autonomous drone platform integrating thermal, RGB, and LiDAR sensing for wildfire detection and terrain mapping.
- Deployed YOLOv8-S and thermal CNN heads on NVIDIA Jetson Orin Nano using ROS2 fusion for low-latency onboard inference.
- Implemented visual-inertial EKF localization and Bayesian fusion for precise georeferencing and flame discrimination.
- Developed an active learning loop with semi-supervised retraining and OTA updates to improve detection across environments.

TraceCloud

May 2024 – Sept 2024

Software Engineer Intern

San Jose, CA

- Developed a scalable full-stack QA automation framework using Selenium WebDriver, Node.js, and PostgreSQL.
- Integrated REST APIs with continuous regression pipelines in Python and Jenkins, reducing test runtime by 40%.
- Engineered a GraphRAG pipeline that modeled requirements, dependencies, and Jira artifacts as graph embeddings, enabling low-latency semantic retrieval and cutting manual trace analysis by 30%.

Projects

ViewLingo | Vapi, OpenAI Realtime API, ElevenLabs, Whisper, React Native, Mentra OS

2025

- Built a real-time AR translator using smart glasses with Whisper transcription and Vapi-powered conversational learning.
- Developed an agentic memory pipeline that tracks user context and adapts next-day lessons based on learning history.

SentraAI | DeepL, Perplexity Sonar Pro, ElevenLabs, Apify, MCP, FastAPI

2025

- Developed a multilingual emergency response agent converting unstructured speech into structured MCP dispatch actions.
- Integrated DeepL translation, Perplexity classification, and Apify retrieval for location-aware autonomous routing.

MedWise | Next.js, Flask, Neo4j, Pinecone, Hugging Face

2025

- Developed a graph-based RAG system linking 1,000+ drugs with embeddings in Pinecone for generics and interaction queries.
- Integrated Neo4j graph traversal with Hugging Face rerankers to improve retrieval precision and clinical interpretability.

Optifi | React, Next.js, FastAPI, PostgreSQL, Docker

2025

- Built a full-stack ERP with autonomous agents orchestrating invoicing, inventory, and workflow automation pipelines.
- Implemented FastAPI microservices with PostgreSQL event logs and Dockerized task runners for real-time execution.