ANISH KAMATAM

San Francisco Bay Area • (925) 549-0603 • anishkamatam22@gmail.com • linkedin.com/in/anishkamatam

Education

California State University - Chico   **Expected May 2027**
B.S. in Computer Science and Mathematics   *Chico, CA*
- **Coursework:** Operating Systems, Computer Networking, Theory of Computation, Data Structures & Algorithms, Robotics & Machine Intelligence, Applied Machine Learning, Artificial Intelligence, Applied Linear Algebra, Programming & Algorithms I & II

Skills

Python, C++, Go, Rust, Java, TypeScript, SQL, React, Node.js, FastAPI, PyTorch, TensorFlow, Scikit-Learn, TensorRT, Docker, Kubernetes, gRPC, Redis, PostgreSQL, AWS, GCP (Vertex AI), LangChain, LangGraph, Vector Databases

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 Krentsel on SLO-driven scheduling for scalable cloud autonomy.

FoodHealth Company   **Nov 2025 − Present**
AI/ML Engineer Intern   *San Francisco, CA*
- Trained XGBoost models on large tabular datasets, optimized hyperparameters with Optuna, reducing inference latency by 30%.
- Built a quantile-scaled, target-encoded feature pipeline with polynomial interactions, improving model accuracy by 12%.
- Developed an audit suite using SHAP, 5-fold CV, and KS-drift checks, increasing model reliability and drift detection by 25%.

FundMore AI   **May 2025 − Oct 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, ElevenLabs, React Native, Mentra OS* — **YCombinator x Mentra Hackathon Winner**   **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* — **AWS MCP x A2A Hackathon Winner**   **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* — **SF Hacks 2x Category Winner**   **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.

Wizard | JavaScript, Node.js, TypeScript, Docker, Postgres, Shell scripting, JSON parsing   **2025**
- Built Wizard, a CLI debugging tool with sandboxed error reproduction, syscall tracing, and structured root-cause analysis.
- Added automated fix workflows using semantic error classification, environment diffing, and safe command execution.