Sanjay Marathe

U.S Citizen | sanjaymarathe.me | maraths1@uci.edu | LinkedIn: sanjay-marathe-2257811b0 | GitHub: SanjayMarathe

Education

University of California, Irvine — Irvine, CA

Bachelors of Science in Computer Science and Intelligent Systems — *Expected Graduation, June 2027*

- **Organizations:** UROP Research Fellow, Association for Computing Machinery, Cyber @ UCI, Product @ UCI

Experience

Geodo.ai — San Fransisco, CA

Software Engineer Intern — *Jan 2026 – Present*

- Engineered a Large Action Model using **Electron** and **Claude Vision** to synthesize desktop demonstrations into autonomous agents with **98%** execution accuracy, reducing manual workflow time by **85%**
- Used locally-hosted vision models and vector databases to filter and index screen recordings, accomplishing autonomous identification of high-signal data and resulting in **100%** reduction in cloud API costs.
- Built an LLM-driven context agent that processes live meeting audio and generate actionable summaries, resulting in a **90%** reduction in manual note-taking time and a **30%** improvement in team asynchronous workflow efficiency

Grangou — Los Angeles, CA

Software Engineer Intern — *Jan 2026 – Present*

- Engineered a high-performance mobile rendering engine for **18,000+** global users using React Native Reanimated, resulting in **60 FPS** scroll performance and a **40%** reduction in memory usage for heavy image feeds.
- Used vector embeddings and cosine similarity algorithms to engineer the platform's core AI matchmaker, resulting in a **92%** match acceptance rate and a **35%** increase in repeat dining bookings.
- Used load-balanced cloud infrastructure to manage a high-traffic product launch, accomplishing the seamless onboarding of **95%** of an **18,000+** user waitlist, resulting in **99.9%** system uptime during peak demand

Projects

DrystAI (Finalist (Top 10%) @ Gemini x Cerebral Valley (Shack15) | *Next, FastAPI, MongoDB Atlas, OpenCV*

- Used DeepGram and Gemini AI to transcribe real-time conversations and generate **3,072-dimensional** embeddings stored in MongoDB Atlas, achieving **85**% accurate semantic retrieval and linking context to recognized faces
- Used OpenCV with FastAPI/Uvicorn to process live video frames for speaker detection and facial recognition, delivering **90%** classification accuracy with smooth real-time streaming to a Next.js frontend
- Used Gemini AI embeddings in MongoDB Atlas Vector Search to perform cosine similarity retrieval across **5,000+** embeddings, achieving **90%** top-1 accuracy and **sub-100ms** query latency for RAG-enabled applications

Boredtutors.com | *React, Node, MySQL, AWS Lambda*

- Developed a full-stack district-wide tutoring platform connecting **250+ students** with **50+ peer tutors**.
- Implemented secure RESTful APIs with JWT authentication, database encryption, and an admin dashboard; later automated tutor verification workflows, cutting manual approval time by **75%**.
- Deployed a scalable cloud architecture on Kubernetes with automated load balancing and failover, ensuring **97.9%** uptime and seamless handling of concurrent user traffic.

MeDisha (1st Place Overall — AWS Agentic AI Hackathon (2025)) | *Next, AWS Bedrock, Pinecone*

- Used Agentic AI and AWS Bedrock vector storage to revolutionize clinical trial matching, resulting in fast, accurate patient-to-trial matching based on comprehensive medical profiles and simplified eligibility criteria
- Used algorithmic ranking to optimize trial site selection, resulting in smart location recommendations based on availability and transparent cost estimations for out-of-pocket versus sponsor-covered expenses.

Technical Skills

Languages: Java, Python, JavaScript (ES6+), SQL, Go, C++
Frameworks & Platforms: React, Node.js, Express, AWS Lambda, Django REST Framework, FastAPI
Databases: PostgreSQL, MySQL, Prisma, Sequelize, Redis
AI/ML & Libraries:Langchain, CopilotKit, NumPy, Mastra, LangGraph