

Pranav Sethia

pranav.sethia5@gmail.com | (949)-326-8956 | LinkedIn: [pranavsethia](#) | GitHub: [pranav-sethia](#)

EDUCATION

University of California, Irvine

Expected Graduation: June 2028

B.S. in Computer Science

GPA: 3.9

Related Coursework: Data Structures & Algorithms, Machine Learning, Discrete Mathematics, Statistics, Linear Algebra

SKILLS

Languages: Python, JavaScript, TypeScript, C++, SQL, HTML/CSS, Bash

Frameworks: React, Next.js, Node.js, FastAPI, PyTorch, LangChain, LangGraph, Docker, AWS, Azure, GitHub Actions

Libraries & Tools: Git, OpenCV, MediaPipe, scikit-learn, Pandas, NumPy, PostgreSQL, pgvector, ONNX Runtime, Gemini API

EXPERIENCE

Humyn Labs ([humynlabs.ai](#)) - supplying egocentric video for robotics AI training

San Francisco, CA

Machine Learning Engineer Intern

June 2026 - Present

- Built and shipped a production PII-blurring pipeline to OpenAI - a 4-model ensemble (SCRFD, YOLOv8, RetinaFace) with ByteTrack tracking - hitting 91.5% recall at 0 false positives on a self-built benchmark
- Deployed a Gemini 2.5 video-QA pipeline on AWS (S3 + EC2 GPU) that flags staged tasks, fake pacing, and idling before OpenAI delivery, using prompt engineering and self-consistency sampling to reach 93-97% reliability
- Engineered a video scene-classifier spanning 1,092 location labels (warehouse, workshop, retail), pairing Gemini with embedding search to solve accuracy drops when the model evaluates over 100 choices at once

Acceler8 ([useacceler8.com](#)) - a16z seed funded startup developing talent solutions

Singapore

Software Engineering Intern

June 2025 - Sep 2025

- Cut client review time from 3 days to <1 day (40% faster) by building an automated pipeline that processes 1,000+ AI summaries and formats them into stakeholder-ready Excel reports using LangChain, Pandas, and OpenPyXL
- Built a synthetic data generator using LLMs and Python scripts to create 1,000+ diverse, realistic employee personas, enabling automated regression testing that caught 8 critical NLP bugs before production deployment
- Implemented an end-to-end Pytest framework validating 50+ FastAPI endpoints across authentication, data processing, and reporting services, catching 12 critical bugs pre-production and reducing QA time by 30 hours/month

PROJECTS

HackSwipe ([hackswipe.com](#)) - Personalized news recommendation engine

Python, TypeScript, React, Node.js, PostgreSQL (pgvector), Transformers.js, Azure, GitHub Actions, Hugging Face Spaces 2026

- Architected a recommendation engine using pgvector cosine similarity across 384-D embeddings, layering recency decay and MMR diversity filtering to reserve 20% of every feed for discovery content
- Automated the app's content pipeline via a GitHub Actions cron job that ingests and summarizes 50 articles every 2 hours using Llama 3.1 8B (Groq) under a strict JSON schema, with backoff retry on rate limits
- Cut signup friction with an ephemeral guest JWT system that instantly migrates a user's full swipe history and taste profile (personalized preferences) to a permanent account via SQL reassignment on Google sign-in

ADYou ([Won 2nd Place, IrvineHacks](#)) - Assesses feasibility of building ADU using AI agents

Python, TypeScript, Next.js, FastAPI, PyTorch, LangGraph, ChromaDB, LangSAM, OpenCV, Mapbox, Turf.js, Shapely 2026

- Designed a multi-agent Retrieval-Augmented Generation (RAG) system with LangGraph and ChromaDB to query 500-page regulatory corpus, implemented adaptive LLM routing and SSE token streaming to achieve <2s latency
- Engineered a scalable computer vision API with FastAPI and PyTorch for satellite object detection, using OpenCV and Shapely morphological operations to transform raw pixel tensors into sub-meter accurate geospatial polygons
- Developed a full-stack Next.js and TypeScript spatial editor using Mapbox GL, optimizing system architecture by offloading complex geometric computations to the client via Turf.js to enable zero-latency, serverless rendering

Project Lead (AI@UCI) - StudyWise & Agentic AI workshops

Python, FastAPI, React, OpenCV, scikit-learn, WebSockets, MediaPipe, Vite, Claude, Cursor

2025 - 2026

- Led end-to-end development of StudyWise, achieving 96% accuracy by training a scikit-learn Random Forest on 30,000+ labeled frames using 5-fold cross-validation and landmark feature extraction via OpenCV and MediaPipe
- Architected a FastAPI WebSocket server with rolling-window signal processing to stream posture data at 15 FPS to a React dashboard, enabling real-time focus analytics with zero-latency feedback
- Delivered 3 Agentic AI workshops to 50+ students on multi-agent systems and LLM architecture