

Pranav Kumar Soma

MACHINE LEARNING ENGINEER • AI RESEARCH

prsama@ucsd.edu • LinkedIn • pranavsoma.me • San Diego / SF Bay Area

EDUCATION

M.S. Computer Science, UC San Diego

Dec 2026 (expected)

GPA 4.00 • Jacobs School of Engineering

B.S. Computer Science, UC San Diego

Dec 2025

GPA 3.96 • Cum Laude, CSE Department Honors, Provost Honors (all quarters)

Coursework: Data Systems for ML, Theoretical Aspects of LLMs, Statistical NLP, Search & Optimization, ML: Learning Algorithms, Computer Vision, Recommender Systems & Web Mining, Convex/Discrete Optimization, Design & Analysis of Algorithms.

EXPERIENCE

AI Research Fellow & Agentic Systems Lead | Laboratory for Emerging Intelligence, UCSD

Apr 2025–Present

- › Fine-tuned model-agnostic tutoring LLMs (**LoRA + SFT + RLHF**) with **Bayesian Knowledge Tracing**; deployed AI Tutor to 3 universities for **8,000 students / 8 courses** • **64% fewer TA hours, 87% higher engagement** (A/B tested across assignments)
- › Lead **Teradata-funded** agentic-AI research • custom **MCP servers** over UCSD APIs, automated skill generation & optimization, interpretability of user intent vs. skill instructions; MongoDB vector store, FastAPI serving

Graduate Researcher | Qualcomm Institute / Calit2 (HXI Lab)

Nov 2024–Dec 2025

- › Built **multimodal** ML pipelines over **TB-scale** sensor/audio/video/image data, parsing patient signals into structured health-database queries for the LifeSaver AI assistant (emergency services, elderly, rural care)
- › Fine-tuned locally hosted **Qwen** models (SFT) on the SD Supercomputer Center; built **GraphRAG** over the UMLS medical ontology • **EMS-certified** diagnostic accuracy at emergency-grade latency
- › Engineered the robotics pipeline tying data queries to motor/servo actuation; first-author lab poster

Machine Learning Researcher (Honors Thesis) | UCSD CSE • Computational Redistricting

Sep 2024–Present

- › Trained per-target **XGBoost** regressors on per-block geospatial/demographic features + **multilevel graph partitioning** to generate districting plans (pooled fav-share **MAE 0.042** across 9 states, held-out cross-state generalization)
- › Implemented a **Sequential Monte Carlo** sampler (spanning-tree balanced cuts, N=128 particles) and greedy DP baselines; authored VRA Section-2 / Polsby–Popper / efficiency-gap fairness metrics. **Applying to ACM CHI 2026**; presenting at the SF Exploratorium, Dec 2026

Software Engineer Intern | ServiceNow

Jun–Sep 2025

AI Voice Agents • **Technical Skills Award** • top of 800+ interns

- › Re-architected the voice-agent **NLU** pipeline onto higher-capacity base models across 8 intents • **75% lower inference latency**; shipped to enterprise customers (Disney+). **Return offer received**

SELECTED PROJECTS

- › **ORCA • AI Music Production** (*Best Project, SDTC: 1 of 30 from 200+; SAIRS 2025*) • hybrid neural networks for source separation, CNN audio-to-score, OpenCV vision; **98% arrangement accuracy** (SME, layman & cross-entropy validated)
- › **BioVeritas** • AI platform evaluating logical rigor & authenticity in biological research via retrieval/context systems
- › **Skin Lesion Detector** • embedded image-classification system for dermatological lesion identification (TensorFlow, OpenCV)

LEADERSHIP

Co-Founder & President, Real World AI Network (2026–Present) • founded **7,000+ student** federation across UCSD's engineering, data-science & business schools driving AI research & entrepreneurship.

Founding Board Member & VP External Affairs, University AI Alliance • built nation's largest student-led frontier-tech network (MIT, Stanford, Caltech, Cornell); backed by a16z speedrun, Bain Capital Ventures, Techstars.

President, CSE Society @ UCSD (100+ members) • **Founder & President**, E/Acc Research Group • **Awards:** National Merit Finalist, USACO Silver, AP Scholar w/ Distinction, YC Startup School 2026.

TECHNICAL SKILLS

ML & AI: PyTorch, TensorFlow, scikit-learn, XGBoost, OpenCV; LLM fine-tuning (LoRA, SFT, RLHF), RAG & GraphRAG, LLM agents & MCP, Transformers, CNNs, RNNs, LSTMs, reinforcement learning, NLP/NLU, Bayesian Knowledge Tracing, Monte Carlo Tree Search, Sequential Monte Carlo, recommender systems, computer vision; Ollama, Qwen, Amazon Lex, Now Assist

Languages: Python, Java, C++, C, JavaScript, TypeScript, SQL, Swift, Kotlin, ARM-32 Assembly, Bash, HTML/CSS

Data & HPC: NumPy, Pandas, Matplotlib, GeoPandas, Shapely, pymetis, Parquet/PyArrow, parallel & multi-threaded computing, SD Supercomputer Center, spatial databases

Infra & Tooling: FastAPI, Flask, Django, React, Node/Express; MongoDB (vector), PostgreSQL, MySQL; Docker, AWS (Lex, Lambda, CloudFormation), Git/GitHub, Linux, Jupyter, JUnit, LaTeX, Microsoft Office, Google Workspace