

Pranav Kumar Soma

SOFTWARE ENGINEER • MACHINE LEARNING • AI RESEARCH

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

EDUCATION

M.S. Computer Science, UC San Diego

GPA 4.00 • Jacobs School of Engineering

Dec 2026 (expected)

B.S. Computer Science, UC San Diego

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

Dec 2025

Coursework: Data Systems for ML, Theory of LLMs, Statistical NLP, Search & Optimization, Computer Vision, ML Algorithms, Recommender Systems, Algorithms, Operating Systems, Theory of Computation.

EXPERIENCE

Software Engineer Intern | ServiceNow

AI Voice Agents, IT Service Management • **Technical Skills Award** • top of 800+ interns

Jun–Sep 2025

- Migrated AI Voice Agents platform from deprecating AWS Lex V1 to V2 ahead of schedule; shipped to production, validated live with Disney+
- Re-architected NLU pipeline onto higher-capacity models across 8 intents • **75% lower response latency**; built dynamic post-incident “wrap-up” modal (UI + backend). **Return offer received**

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

Apr 2025–Present

- Lead **Teradata-funded** agentic-AI platform for 7,000+ users • custom **MCP servers** over UCSD APIs + deterministic, fault-graceful workflow executor
- Built Python **AI Tutor** (UCSD, SDSU, Cal Poly Pomona) for **8,000 students / 8 courses** • **64% fewer TA hours, 87% higher engagement**
- Fine-tuned model-agnostic LLMs (**LoRA + SFT + RLHF**) w/ Bayesian Knowledge Tracing, MongoDB vectors, FastAPI serving

Graduate Researcher | Qualcomm Institute / Calit2 (HXI Lab)

Nov 2024–Dec 2025

- Built multimodal ML pipelines over **TB-scale** sensor/audio/video, parsing patient data into structured health-database queries (LifeSaver AI assistant)
- Fine-tuned local Qwen models (SFT) on SD Supercomputer + **GraphRAG** over UMLS ontology • **EMS-certified** diagnostic accuracy at emergency-grade latency

Honors Thesis Researcher | UCSD CSE • Computational Redistricting

Sep 2024–Present

- Built geospatial ML pipeline scoring U.S. redistricting plans over Census-2020 data (**158K–5.5M blocks/state**)
- Engineered 3 generators (greedy DP + Kernighan–Lin, Sequential Monte Carlo, **XGBoost + graph partitioning**, MAE 0.042); authored VRA / Polsby–Popper fairness metrics. **Applying to ACM CHI 2026**; SF Exploratorium, Dec 2026

SELECTED PROJECTS

- **ORCA • AI Music Production** (*Best Project, SDTC: 1 of 30 from 200+; SAIRS 2025*) • OpenCV + hybrid neural nets for source separation, CNN audio-to-score; **98% arrangement accuracy**. Flask, Docker, React
- **Spay.LA** (*spay.la*) • MERN fundraising & legislative platform; **raised \$75K**; led 11 developers. Firebase/Google auth
- **Union Station Homeless Services** • React listing platform w/ RegEx validation & state machines, SheetJS/MongoDB bulk exports for LA’s sole housing provider

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 • **Engineering Manager**, Triton Software Engineering.

TECHNICAL SKILLS

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

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

Web, Data & Cloud: React, Node.js, Express, FastAPI, Flask, Django/DRF, MERN, React Spring, SheetJS, Firebase, Google OAuth, REST APIs; NumPy, Pandas, Matplotlib, GeoPandas, Shapely, pymetis, Parquet; MongoDB (vector), PostgreSQL, MySQL; AWS (Lex, Lambda, CloudFormation), Docker, Git/GitHub, Maven, Gradle, Linux, parallel computing, SD Supercomputer Center

Robotics, Enterprise & Tools: ROS, WPIlib, PID control, finite state machines, embedded systems; ServiceNow (ITSM, CSM, SPM), UI Builder, Smartsheet, Windsurf, Android Studio, Xcode, Jupyter, JUnit, RegEx, Agile/Scrum, system design, LaTeX, Microsoft Office (Word, Excel, PowerPoint), Google Workspace