

Pravir S. Chugh

(425) 445-2056 | pravirchugh@ucla.edu | www.linkedin.com/in/pravirchugh/

EDUCATION

University of California, Los Angeles
B.S. in Computer Science

Expected Graduation: June 2026
GPA: 3.79/4.00

Relevant Coursework: Software Engineering · Machine Learning · Natural Language Processing · Data Mining · Operating Systems · Computer Systems Architecture · Logic Design · Computer Organization · Algorithms

TECHNICAL SKILLS

Languages: Python · C · C++ · Java · JavaScript · SQL

AI / ML: PyTorch · TensorFlow · Hugging Face · Model Evaluation · Embedding Models · RAG · Agents

Systems / Networks: Linux · Distributed Systems · Multi-threading · Concurrency · Containers (Docker) · Git

Cloud / Databases: AWS · SAP BTP (Cloud Foundry) · CI/CD (GitHub Actions) · SAP HANA · MongoDB

WORK EXPERIENCE

SAP **June 2024 – Present**

AI/ML Intern

- Designed a scalable, embedding-based retrieval engine for large Java codebases, enabling modular chunking, indexing, and querying, and contributed to a pending patent based on this design.
- Improved throughput of internal LLM endpoints by introducing multi-threading and reducing unnecessary locking in Python/C++ service components, increasing sustained request rate from ~15 to ~100 req/min.
- Built the end-to-end onboarding and orchestration pipeline for Ariba's recommendation system, including customer enrollment flows, secure data ingestion, scheduled model training jobs, and automated data deletion for opt-outs.
- Replaced legacy keyword search with a semantic vector search system (embedding models with SAP HANA), improving product match relevance to ~90% recall across procurement categories.
- Added code summarization to Ariba's review workflow, cutting reviews from ~5 days to ~2 days for 50+ engineers.
- Led systems architecture discussions and code reviews across platform and ML teams to ensure reliability, correctness, and safe rollout of new components.

UCLA Security Lab

November 2024 – June 2025

Research Assistant

- Developed a multi-agent LLM framework to detect vulnerabilities in binary programs by analyzing execution paths and memory behavior, enabling automated CWE classification without access to source code.
- Engineered and evaluated model performance on real-world and synthetic C program datasets, achieving 92% accuracy in multi-class CWE classification and co-authoring a paper submitted to ACM CCS 2025.

Bruin AI @ UCLA

October 2023 – June 2025

Director of External Affairs

- Secured partnerships with AI industry leaders, creating internship and project opportunities for club members.
- Delivered technical presentations on LLM training and fine-tuning workflows to ~40 students.

Stanford Institute for Human-Centered AI

June 2023 – September 2023

AI Research Intern

- Fine-tuned transformer models on GPT-4/LLaMA outputs to analyze how biases propagate in downstream models.
- Evaluated ~100K model-generated outputs and quantified bias using embedding-based association metrics, identifying a 15% amplification in downstream tasks.