

KUNJ SHAH

✉ kunjcr2@gmail.com

in [linkedin.com/in/kunjcr2](https://www.linkedin.com/in/kunjcr2)

github github.com/kunjcr2

globe kunjcr2.github.io

Education

San Jose State University

May 2027

Bachelor of Science in Computer Science

GPA: 3.88/4.00, Dean's List

Coursework: Object Oriented Programming, Advanced DSA, Computer Architecture, Information Security, Machine Learning, Software Engineering

Experience

A10 Networks

May 2026 – Aug 2026

Deep Learning Engineer Intern

San Jose, CA

- Architected and trained **TrooperViz** to support multimodal safety evaluation across text, image-text, multilingual, and multi-turn inputs, scaling PyTorch training across **8 GPUs** and **180k examples**.
- Fine-tuned a vision-language model for safety classification and merged it with a text-only guardrail using **SLERP**, creating a unified model for text, image, conversational, and generative-AI safety evaluation.
- Built end-to-end data and evaluation pipelines with **Pandas** and **Polars**, combining synthetic-data generation and automated benchmarking with interactive inference through **Gradio** and **vLLM**.

Routes Technologies

Oct 2025 – Present

AI Engineer Intern

Remote

- Architected an LLM-powered **NL-to-SQL** pipeline using few-shot prompting, input normalization, **470+** synonym mappings, parameterized SQL sanitization, session memory, and **Weights & Biases** observability, deploying the system across **5 Azure ML endpoints**.
- Built an **SVD collaborative-filtering** recommendation system with scikit-learn and an LLM-based taste profiler, generating updated user preference profiles through automated nightly ETL jobs.
- Engineered a multi-source recipe ingestion pipeline using **Scrapy**, social-media APIs, LLM classification, language filtering, and **Faster-Whisper** transcription to process video and web-based content.

Research under Prof. Mark Stamp

Oct 2026 – Present

Adversarial Neural Networks for Cryptography

San Jose State University

- Investigating how neural-network architecture affects attacker performance in adversarial neural cryptography by comparing architectures across controlled cryptographic experiments.
- Implemented and evaluated attacker architectures in **PyTorch**, analyzing convergence, key-recovery behavior, and cross-architecture generalization to characterize architecture-dependent attack performance.

Dreamable Inc.

May 2025 – Aug 2025

AI/ML Engineering Intern

San Francisco, CA

- Fine-tuned a **Qwen 7B** model using TensorFlow, Hugging Face, and Lambda, applying **LoRA** to reduce training cost and memory usage while delivering a domain-specific question-answering model.
- Evaluated model configurations across multiple hyperparameters through systematic experimentation, tracking training and evaluation metrics with **Weights & Biases** to optimize model performance.
- Developed an AI outreach agent using **LangChain**, **Exa.ai**, and LLM APIs to automate messaging workflows, reducing manual outreach time and improving response rates while documenting processes in Notion.

Leadership / Extracurricular

AI Club at SFSU, Vice President

Jun 2025 – Dec 2025

SparkSF, Tech Lead

Jun 2024 – Dec 2025

Duc Ta, CS215, Teaching Assistant

Jan 2024 – Dec 2024

CodePath, Advanced DSA Batch

Jan 2026 – May 2026

Technical Skills

Programming: Python, SQL, Java, JavaScript, Git

Machine Learning & NLP: PyTorch, TensorFlow, scikit-learn, Weights & Biases, Pandas, NumPy, Neural Networks, Reinforcement Learning, LangChain

LLMs: Transformers, Pretraining, Fine-tuning, Inference, vLLM, Alignment, Benchmarking, Multi-GPU Training

Backend & Cloud: Flask, FastAPI, Node.js, PostgreSQL, MySQL, Docker, CI/CD, Microsoft Azure, Lambda