Mira Saini | AI & ML Engineer | Santa Cruz, Ca

misaini@ucsc.edu | (916) 412-7952 | linkedin.com/in/mira-saini | simba4077.github.io/personal_site/

Education

University of California, Santa Cruz: B.S. Computer Science 2027

Relevant Coursework: Algorithms & Data Structures, Neural Networks, Data Visualization, ML, AI

Experience

UC Santa Cruz Tech4Good Research Lab — Generative AI Lead Researcher **Dec 2024 – Present**

- Directed a 6-member engineering team (with 2 PhD mentors) to develop an OpenWebUI-based UCSC community chatbot used by more than 300 students; integrated Pinecone (llama-text-embed-v2) and engineered CSV/JSON schemas and chunking strategies to elevate retrieval quality.
- Architected Python tool-calling and routing modules that deliver semantic club/event retrieval and contextual recommendations; maintained more than 200 structured knowledge-base records.
- Devised an LLM-driven auto-invite workflow that interprets dialogue context to place users into more than 5 topic-specific channels, improving relevance and user engagement.
- Analyzed more than 1,000 chat logs to iteratively refine prompt structure and retrieval logic, decreasing hallucinations and strengthening grounding during testing.

Projects

ATS Resume Optimizer | Python, FastAPI, OpenAI API, React/Vite

Github | Demo

- Increased ATS keyword match rates by ~50% using automated JD parsing and targeted resume rewriting.
- Built PDF/DOCX ingestion and reconstruction pipeline that cut resume tailoring time from ~20 min to less than 1 min.
- Improved formatting consistency with structured extraction and regex-based cleanup, reducing output errors by ~70%.

AddIt | Python, Streamlit, Gemini API, Flask

Github

- Constructed an AI workflow to ingest event posters which extracts structured fields (title, time, location) and auto-generates Google Calendar events, cutting manual entry time by ~80%.
- Attained ~95% field extraction accuracy across ~50 posters; incorporated a Streamlit UI for user editing.
- Enhanced reliability with defensive parsing and error-handling for OCR/LLM edge cases.

CNN CIFAR-10 Image Classifier | Python, NumPy

Github

- Implemented a CNN using NumPy; assembled full train/val/test pipeline with reusable training utilities.
- Reached ~60% accuracy on CIFAR-10; tracked convergence using structured loss and metric reporting.

MLP MNIST Classifier Training Pipeline | Python, NumPy

Github

- Wrote a minimal MLP training loop (train/val split, batching, accuracy/loss reporting) with dataset preprocessing to classify handwritten digits in MNIST dataset with accuracy of more than 97%.
- Rapid hyper-parameter tuning using real-time train/val metric reporting, leading to stable convergence of model.

Skills

Programming Languages & Tools: Python (pandas, NumPy, PyTorch, Tensorflow, regex), C, SQL, Git/GitHub, GCP, Firebase

AI & Agentic Systems: RAG pipelines, agent workflows, embeddings (Pinecone), MCP servers, prompting, evals

Data Engineering: JSON/CSV schema design, data cleaning/normalization, logging, labeling, MySQL, MongoDB