

Mira Saini

AI Engineer | Santa Cruz, CA

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

TECHNICAL SKILLS

Machine Learning: PyTorch, TensorFlow/Keras, scikit-learn, MLPs/CNNs, model training/evaluation

Data Engineering: Python (NumPy, Pandas), ETL, dataset cleaning, feature extraction

LLMs & RAG: Pinecone, embeddings, OpenWebUI, prompt engineering

Backend: FastAPI, Flask, REST APIs, Streamlit

Tools: Git/GitHub, Docker, Linux, CI/CD (GitHub Actions)

Foundations: Neural networks, optimization, evaluation metrics, DSA

EDUCATION

University of California, Santa Cruz

Santa Cruz, CA

B.S. Computer Science

June 2027

- GPA: 3.9/4.0; Coursework: Machine Learning, Computer Systems, Discrete Math, Data Structures and Algorithms

EXPERIENCE

Tech4Good Lab

Dec. 2024 – Present

Generative AI Lead Researcher & Software Engineer

Santa Cruz, CA

- Lead a 6-member agile engineering team, owning end-to-end development of an OpenWebUI-based UCSC community chatbot used by more than 300 students.
- Designed CSV/JSON-based data models and ingestion pipelines integrated with Pinecone, improving retrieval reliability and accuracy across club and event knowledge.
- Implemented an LLM-driven auto-invite workflow that transforms dialogue context into structured routing decisions, increasing engagement across more than 5 topic-based channels.
- Analyzed over 1,000 chat logs with Python to identify common LLM failure modes and applied targeted “patches” to improve response trajectories, grounding, and tool-use consistency.

PROJECTS

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

- Built a full-stack system that parses job descriptions and rewrites resumes using LLM-based pipelines.
- Implemented PDF/DOCX preprocessing and structured JD parsing, reducing tailoring time by 90%.
- Improved ATS keyword alignment using automated skill extraction and structured outputs.

Multi-Genre Music Classification | Python, Pandas, NumPy, Tensorflow

- Created a 7,200 sample audio dataset with 115 extracted spectral and temporal features using Librosa.
- Developed preprocessing and feature-extraction scripts; cleaned the dataset down to 6,552 reliable samples.
- Implemented a baseline MLP classifier (NumPy/PyTorch) to evaluate dataset quality (around 66% accuracy).
- Benchmarked model performance across variations in hyperparameters, data cleaning strategies, and preprocessing pipelines to identify failure modes and optimize training efficiency.

AddIt | Python, Streamlit, Gemini API

- Developed an app that extracts event metadata (title, time, location) from posters using multimodal prompting.
- Achieved around 5% extraction accuracy across more than 50 posters using validation checks and OCR.
- Built a user-review interface for correcting low-confidence outputs.

Neural Networks From Scratch (CNN & MLP) | Python, NumPy

- Implemented MLP (97% MNIST) and CNN (60% CIFAR-10) from scratch with full backprop and batching.
- Analyzed training curves, overfitting, and optimizer behavior across experiments.