

Sofi Altamsh Shah *Machine Learning Engineer*

✉ sofialtamsh123@gmail.com ☎ +91-9156704982 📍 Bengaluru, Karnataka 📄 Portfolio

🌐 linkedin.com/in/sofialtamsh 🏠 github.com/sofialtamsh

Professional Summary

Machine Learning Engineer with 2+ years experience in LLMs, speech/audio ML using Whisper/torchaudio, and RAG. Skilled in fine-tuning models, optimizing inference, and building scalable ASR/TTS pipelines. Delivered production ML systems, reducing QnA response time by 92% and improved hiring rate by 23%.

Experience

Machine Learning Engineer, Masai 01/2025 – present | Bengaluru

- Built an automated **multilingual audio-generation** pipeline that converted recorded sessions into five languages using ASR, translation, and TTS models, reducing manual effort by over **80%**.
- Designed and built multi-stage **RAG + AI-agent pipelines** using LangChain for intent extraction, context retrieval, and automated support ticket responses.
- Built a RAG-based answering system for recorded sessions, reducing query time by **92% ($\leq 3.2s$)**. Improved student hiring rate by **23%** using a CTC prediction model. Trained and deployed to production.

Data Analyst Associate → Data Research Analyst, Masai 04/2024 – 01/2025 | Bengaluru

- Build data pipelines by analyzing the job market data from (LinkedIn, Naukri, Instahyre), delivering insights that improved placement rates by **15%**. Analyzed trends and built dashboards.

Data Science Intern, IStop.ai 06/2022 – 11/2022 | Bengaluru

- Trained predictive models for house-loan value achieving more than **90% accuracy**, enhanced performance using **PCA**, and built insight dashboards with clustering visualizations.

Technical Skills

Audio & Speech ML: Whisper, torchaudio, librosa, MFCCs, VAD, augmentation, noise filtering

Machine Learning: Regression, Clustering, PCA, XGBoost, Feature Engineering, Hyperparameter Tuning

Deep Learning: CNN, RNN, GANs, Transformers (RoBERTa, T5, GPT), LoRA

NLP & LLMs: Prompt Engineering, RAG, Text2Text, NER, NLTK, Hugging Face, Fine-Tuning, LangChain

MLOps: Streamlit, Flask, HDFS, AWS Lambda, Vector Databases (FAISS), MySQL, MongoDB, Docker

Libraries & Frameworks: NumPy, Pandas, TensorFlow, PyTorch, Scikit-learn, SciPy, OpenCV, OCR

Visualization: Power BI, Matplotlib, Seaborn, Excel, Metabase

Projects

Multilingual Speech-to-Text Enhancement

- Fine-tuned Whisper using **torchaudio preprocessing and audio augmentation** on English–Hindi data
- Achieved **22% WER reduction** and optimized real-time inference with FP16 and caching

Fake News Detection Using Multi-Modal Learning

- Built a hybrid multimodal model (RoBERTa + CNN) for fake news detection, achieving **88.7% accuracy** on a 10K+ sample dataset using PyTorch & Hugging Face.

Face Mask and Emotion Detection System

- Built a **real-time image pipeline** with **94.6% accuracy** on **1,500+ images** using TensorFlow and OpenCV

Education

B.Tech in Computer Science and Engineering,
Government College of Engineering Aurangabad (8.1 CGPA)

2019 – 2023