

Mousami Soni

AI/ML Engineer | Applied ML | NLP & Audio AI

Berlin, Germany | +49 151 5885 2196 | soni.mousami7@gmail.com
linkedin.com/in/mousami7 | github.com/Mousami7 | Portfolio

Professional Summary

AI/ML Engineer with **5.5+** years of experience building **production-grade AI and ML pipelines** and **API-based inference systems** in Python. Strong background in **applied NLP** and **audio intelligence** (speaker diarization, speaker verification, content moderation), with hands-on delivery using **Apache Airflow**, **FastAPI**, **SQL**, and **AWS** (S3, EC2, Lambda, DynamoDB). Experienced in taking systems from prototype to production with a focus on **reliability**, **monitoring**, and **scalable system design**, and actively expanding into modern **LLM-based AI system architectures**.

Core Skills

Programming: Python, SQL

AI/ML: Machine Learning, NLP, Speech & Audio ML, Classification, Model Evaluation

AI Systems: End-to-end ML pipelines, embeddings & vector-based retrieval (conceptual), RAG (foundational)

Frameworks: scikit-learn, TensorFlow, PyTorch, Whisper, SpeechBrain

Data & ETL: Pandas, PostgreSQL, MySQL, DynamoDB

Orchestration: Apache Airflow (DAGs, scheduling, retries, monitoring)

Backend & Cloud: FastAPI, REST APIs, **AWS (S3, EC2, Lambda)**, Docker, Git; strong understanding of cloud fundamentals and comfortable working with **Azure**-based services

Experience

Xounds (EXIST-Funded Startup)

AI/ML Engineer

Berlin, Germany

Dec 2024 – Present

- Designed pipeline state tracking using **AWS DynamoDB** to monitor **50+ audio jobs/day**, improving failure visibility and reducing manual debugging by **40%**.
- Built end-to-end ML workflows for **speaker diarization**, **speaker verification**, and **audio content moderation** using Python-based models.
- Orchestrated multi-stage audio processing pipelines with **Apache Airflow**, including task dependencies, retries, and monitoring.
- Developed **FastAPI**-based inference services integrated with **AWS S3, EC2, and Lambda** for scalable audio analysis.

Ceamless Events

Data/AI Developer

Berlin, Germany

Jun 2024 – Nov 2024

- Supported event operations by managing and organizing attendee and ticketing data for multiple events.
- Built simple data ingestion workflows for CSV and Excel files into PostgreSQL to maintain structured event records.
- Performed data cleaning, normalization, and deduplication to improve the accuracy of attendee information.
- Assisted in generating attendance and ticketing summaries used by organizers for event planning and coordination.

Yess Infotech (Training Institute)

Data Science & Python Trainer

Pune, India

Aug 2022 – Aug 2023

- Delivered hands-on training in **Python**, **SQL**, and core **AI/ML fundamentals**, focusing on practical understanding rather than theory.
- Guided learners through end-to-end data science projects, covering data preprocessing, feature engineering, model training, and evaluation.
- Taught foundational machine learning concepts including classification, regression, and basic model performance metrics.
- Introduced workflow automation and basic deployment concepts using **Apache Airflow** and simple web applications with **Flask**.

Vocational Education Project (MP Govt.)

IT Trainer

India

Sep 2016 – Sep 2019

- Trained **200+** learners in Python, SQL, database concepts, and applied IT skills through projects and assessments.

Projects

Audio Moderation and Diarization Pipeline

Python, Apache Airflow, FastAPI, AWS

- Built an end-to-end audio processing pipeline for speaker diarization, speaker verification, and content moderation.
- Integrated model inference with cloud storage and exposed results via FastAPI for downstream system use.

Event Data ETL and Reporting System

Python, PostgreSQL, Apache Airflow

- Implemented simple data ingestion workflows for event attendee and ticketing data from CSV and Excel sources.
- Generated structured datasets and summary reports to support event planning and coordination.

Master's Thesis: Stock Market Prediction using LLM-based Models

Python, Machine Learning, LLM Concepts

- Conducted a Master's thesis exploring the use of open-source LLMs for predicting stock and cryptocurrency price movements.
- Collected and preprocessed historical market data and compared model behavior across different assets and time windows.

Education

International University of Applied Sciences

M.Sc. Computer Science

Germany

2022 – 2024

Sagar Institute of Research & Technology

M.Tech. Digital Communication

Indore, India

2015 – 2017

Jawaharlal Institute of Technology

B.E. Electronics & Communication

Borawan, India

2010 – 2013

Certifications & Languages

Certifications:

- Apache Airflow (Astronomer) Certification Sep 2025
- Data Analytics & Machine Learning (Puneri Pattern) 2022
- Data Science with Python (Simplilearn) 2021

Languages:

- English (IELTS C1)
- German (A2)