

Jaleed Ahmad

AI/ML Engineer

📍 Faisalabad, Pakistan | 📞 +92 321 1600345 | ✉ Email | 🔗 LinkedIn | 🐙 GitHub | 🌐 Portfolio

Summary

AI/ML Engineer specializing in Deep Learning, NLP, and Generative AI. Proven track record of developing end-to-end intelligent systems, from fine-tuning custom neural networks to architecting scalable LLM pipelines. Combines a strong foundation in model development with expertise in MLOps and AWS cloud infrastructure to deliver robust, production-ready AI solutions.

Skills

Programming Languages: Python

ML/DL Frameworks: PyTorch, TensorFlow, Keras, Transformers, Stable Diffusion

NLP/GenAI: LangChain, Embeddings, Tokenizers, Vector Databases (VDB), LLM APIs

MLOps & Cloud: Linux, Git, GitHub, Docker, DockerHub, Kubernetes, AWS (IAM, ECR, EC2, S3, Lambda)

Work Experience

Trainee AI/ML Engineer | T8 Launchpad (NASTP Delta) – Lahore 01/2024 – Present

- Engineered scalable infrastructure for deploying and serving AI/ML and Computer Vision models using AWS ECS Fargate and Docker.
- Developed and optimized Retrieval-Augmented Generation (RAG) pipelines utilizing Large Language Models (LLMs) to enhance context-aware enterprise workflows.
- Managed serverless data processing pipelines leveraging AWS S3, Lambda, and DynamoDB to support backend AI architectures.
- Orchestrated automated CI/CD pipelines via GitHub Actions and Jenkins to streamline continuous integration and delivery of intelligent applications.

NLP Intern | Elevvo 09/2025 – 10/2025

- Completed an intensive internship focused on Natural Language Processing, executing six diverse text-processing and machine learning projects.
- Developed and fine-tuned NLP models using Python libraries such as NLTK, spaCy, and Hugging Face Transformers.
- Documented methodologies and shared corresponding notebooks in a public repository to demonstrate reproducibility and model performance.

Projects

AI Study Notes Agent | 🐙 GitHub | 🎥 Demo | *Python, Gemini 2.5 Flash, RAG, ChromaDB, Supabase*

- Architected a multi-user, cloud-native AI study assistant leveraging Google Gemini 2.5 Flash and Search Grounding APIs for real-time, context-aware document analysis.
- Engineered a Retrieval-Augmented Generation (RAG) pipeline using LangChain and ChromaDB (gemini-embedding-001), enabling persistent semantic search and accurate citations across large textbook libraries.
- Built automated data extraction pipelines using GenAnki to seamlessly convert unstructured AI notes into structured SQLite-backed spaced-repetition flashcard decks.
- Integrated a PostgreSQL database via Supabase to handle secure user authentication (OAuth 2.0) and persistent chat session management across a multi-tenant architecture.

AI Image Forensics Web Application | 🐙 GitHub | 🎥 Demo | *Python, FastAPI, React, Gemini Vision API*

- Developed an end-to-end application to detect AI-generated images using a Next.js frontend and a FastAPI backend.
- Integrated the multimodal Gemini Vision API to power the core forensic analysis engine and classify manipulated visual data.

ASL Recognition using Deep Learning | *Python, Pytorch, Jetson Nano*

- Designed a real-time sign language recognition system using five MPU6050 sensors to capture complex hand and finger motion data.
- Implemented signal preprocessing, feature extraction, and classification using customized LSTM and CNN architectures.
- Deployed the trained model on a Jetson Nano for low-latency, real-time edge inference.

AuraBeat: Context-Aware Music Recommendation | GitHub | Demo | *Node.js, Google Gemini API, Open-Meteo*

- Built an intelligent web app providing multi-sensory music recommendations through an AI-powered conversational interface.
- Implemented real-time contextual awareness by feeding user mood, live weather, and location data into an LLM context window.

Education

BS Electrical Engineering, Namal University, Mianwali

Certificates

- Oracle Cloud Infrastructure AI Foundations Associate (2025)
- LangChain for LLM Application Development (2025)
- Understanding and Applying Text Embeddings (2025)
- ChatGPT Prompt Engineering for Developers (2025)
- Getting Started with Mistral (2025)

Languages

Urdu: Native | **English:** Professional Working Proficiency