

Abdul Waheed

(AI/ML Engineer - Generative AI)

[in LinkedIn](#) | [Portfolio](#) | [✉ abdulwaheed.aiengineer@gmail.com](mailto:abdulwaheed.aiengineer@gmail.com) | [☎ +92-313-4371205](tel:+92-313-4371205) | [📍 Lahore, Pakistan](#)

Professional Summary

Full Stack AI/ML Engineer with 2+ years of experience in Generative AI, Agentic AI systems, and LLM integration. Skilled in building end-to-end AI solutions like medical assistants, content moderation, and enterprise AI systems. Proficient in MCP, RAG, transformers, and cloud-native deployment, turning research into scalable, production-grade applications. Recognized for driving results, mentoring teams, and contributing to global AI initiatives such as Google Gemini and Twitter AI.

Work Experience

AI/ML Engineer – Generative AI	<u>Hubble42 Inc.</u>	Pakistan 11 /2024 - Current
<ul style="list-style-type: none">Built MedBridge AI, an Agentic AI healthcare assistant using Model Context Protocol (MCP) and GPT-4, automating triage, scheduling, and workflow management.Integrated Google Calendar API and bilingual chat interface, improving patient-doctor matching efficiency and preventing double bookings.Developed XLM-RoBERTa Tweet Classifier for multilingual content moderation (English, Roman Urdu, Urdu, mixed tweets), achieving 82.6% accuracy and 82.4% F1-score, deployed via Flask REST API with real-time predictions.Built a desktop GenAI Interview Assistant powered by GPT-4o, Whisper, and RAG, boosting user engagement by 70% and resume tailoring efficiency by 60%.Designed the frontend using HTML, JavaScript, Tailwind CSS, and containerized with Docker, reducing deployment time by 40%.Developed an AI ERP Assistant using LangChain, FAISS, and LLMs to query structured and unstructured data (SQL, CSV, PDF), cutting enterprise data retrieval time by 50%.Led dataset creation for Google Gemini (GAIA) RLHF/SFT pipelines, generating multi-modal datasets (web, image, video, audio) that supported 20%+ accuracy improvement during fine-tuning.Contributed to Twitter AI (X) by crafting adversarial prompts, performing LLM evaluations, and suggesting completions, improving model alignment and robustness with 15% reduction in hallucinations.Mentored 3+ junior engineers while maintaining delivery quality, improving team code review turnaround by 35%.		

AI/ML Engineer	<u>codSeed</u>	Pakistan 07/2023 - 10/2024
<ul style="list-style-type: none">Built a Text Summarizer by fine-tuning Google Pegasus on the SAMSum dataset, boosting ROUGE-L by 18% and achieving 65% accuracy. Deployed via FastAPI + AWS CI/CD, enabling real-time inference.Automated IT ticketing system using Power Automate and Microsoft Copilot, reducing manual workload by 50% and improving IT support speed by 35%.Developed a Retrieval-Augmented Generation (RAG) chatbot for PDF querying using Streamlit, Milvus (vector DB), and OpenAI API, improving document retrieval accuracy by 40%.Built a Movie Search Agent using LangChain, LLMs, and TMDB API, achieving 90%+ success rate in query understanding and improving user satisfaction.Deployed CI/CD pipelines and cloud workflows, reducing model release cycles by 60% and enabling continuous delivery for 3+ ML apps.Enhanced NLP pipelines with transformer-based APIs, increasing throughput 2.5x and reducing latency by 30%.Engineered a Fake News Detection system, preprocessing and classifying news articles with Logistic Regression, Random Forest, and Decision Tree. Achieved 97% accuracy and 0.95 F1-score, deploying Logistic Regression for production use.		

Education

B.Sc. Computer Science	COMSATS University Islamabad
------------------------	------------------------------

- Majors: Data Structures and Algorithm | Programming | Statistics and Probability | Calculus | Artificial Intelligence | Data Science

Skills

- **Generative & Agentic AI:** GPT-4, Agentic AI Systems, Model Context Protocol (MCP), LangChain, RAG, Autonomous Agents, Prompt Engineering, LLM Fine-tuning (SFT, RLHF), Adversarial Prompting
- **Machine Learning & NLP:** XLM-RoBERTa, MBart, Pegasus, Transformers (Hugging Face, OpenAI), PyTorch, TensorFlow, Scikit-learn, NLP Pipelines (Preprocessing, Feature Engineering, Evaluation)
- **Full-Stack & Deployment:** FastAPI, Flask, Streamlit, REST APIs, Docker, CI/CD (GitHub Actions, AWS CodePipeline), Cloud Platforms (AWS, GCP), Milvus, FAISS, AI-assisted Frontend (React, JavaScript, HTML/CSS with Cursor AI)
- **Data Engineering & Tools:** SQL, MySQL, Pandas, NumPy, Power Automate, Data Visualization (Tableau), Google Calendar API, SMTP Integration, JSON & API-based Data Workflows
- **Soft Skills:** Problem-Solving, Critical Thinking, Communication, Leadership, Collaboration, Mentorship
- **Languages:** English (Professional), Urdu (Native)

Projects

MedBridge AI – Agentic AI Medical Triage & Appointment System:

- **Objective:** Create a healthcare assistant powered by Agentic AI and MCP to handle symptom triage and real-time appointment scheduling.
- **Approach:**
 - Leveraged GPT-4 + MCP integration for intelligent tool calling and autonomous decision-making.
 - Used Google Calendar API for live availability checks, scheduling, and conflict prevention.
 - Added bilingual chat (English/Roman Urdu), speech-to-text, and real-time streaming for enhanced user experience.
- **Impact:** Delivered an end-to-end medical workflow automation system, reducing admin effort, preventing double bookings, and improving patient accessibility through bilingual and voice-enabled features.

XLM-RoBERTa Tweet Classifier – AI-Powered Content Moderation:

- **Objective:** Build a multilingual tweet classification system to detect Hate Speech, Extremism, and Non-Relevant content for safer online spaces.
- **Approach:** Fine-tuned XLM-RoBERTa-base on 3,600+ labeled tweets (English, Urdu, mixed). Implemented preprocessing (URL/mention normalization, text cleaning), deployed via Flask REST API + Hugging Face Transformers, and designed a responsive frontend with real-time feedback.
- **Impact:** Achieved 82.6% test accuracy and 82.4% F1-score, enabling real-time, multilingual content moderation. Improved platform compliance and reduced manual review workload.

GenAI Interview Assistant:

- **Objective:** Build a desktop AI assistant to help users prepare for interviews and tailor resumes effectively.
- **Approach:** Integrated GPT-4o, Whisper, and Retrieval-Augmented Generation (RAG). Designed a modern frontend using HTML, JavaScript, Tailwind CSS, and containerized with Docker for fast deployment.
- **Impact:** Increased user engagement by 70% and improved resume tailoring efficiency by 60%, while reducing deployment time by 40%.

Low Resource English–Urdu Neural Machine Translation (NMT):

- **Objective:** Develop a robust neural machine translation system for English–Urdu in low-resource settings.
- **Approach:** Collected and consolidated 5.7M bilingual sentence pairs, applied preprocessing (HTML tag removal, tokenization, cleaning), and fine-tuned transformer models (MBart, MT5, MarianMT).
- **Impact:** Achieved 35.87 BLEU score using MBart, significantly improving translation accuracy compared to baseline models.

Volunteer

1. Management Head – Calligraphy & Fine Arts Society
2. Event Head – COMSATS Science Club