Abdul Waheed

(AI/ML Engineer - Generative AI)

LinkedIn | Portfolio | abdulwaheed.aiengineer@gmail.com | +92-313-4371205 | Lahore, Pakistan

Professional Summary

AI/ML Engineer with expertise in **Generative AI, Agentic AI systems, and LLM integration**. Skilled in building **end-to-end AI products** including medical assistants, content moderation systems, and enterprise AI solutions. Hands-on experience with **MCP, RAG, transformers, and cloud-native deployment**, turning research into **scalable, production-ready applications**. Recognized for delivering measurable impact, mentoring teams, and contributing to **global AI initiatives** like Google Gemini and Twitter AI.

Work Experience

AI/ML Engineer – Generative AI | **Hubble42 Inc.** | ***Pakistan*** **11 /2024 - Current**

- 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 | **codSeed** | ***Pakistan*** **07/2023 - 10/2024**

- 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.

Vallen AI – Enterprise AI Solutions Website:

- **Objective**: Deliver a modern enterprise website to showcase AI consulting services, industry use cases, and success stories.
- **Approach**: Designed and implemented the site using Cursor AI for AI-assisted development, powered by React 19 + React Router 7. Built modular architecture with 25+ reusable components, responsive UI (Bootstrap 5), and performance optimizations (lazy loading, code splitting). Added SEO, accessibility, and deployment workflows for production readiness.
- **Impact**: Showcased 9 AI services across 6 industries with real ROI-focused case studies (e.g., 80% automation, 15% sales growth). Delivered a professional, scalable platform that strengthened Vallen AI's digital presence and credibility in enterprise markets.

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%.

AI ERP Assistant:

- **Objective**: Enable enterprises to query both structured and unstructured data through natural language.
- **Approach**: Built with LangChain, FAISS, and LLMs, connecting to SQL, CSV, and PDF sources. Designed natural language interfaces for seamless interaction.
- **Impact**: Reduced enterprise data retrieval time by 50%, boosting efficiency for business users.

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.

Text Summarization with Google Pegasus:

- **Objective**: Build a real-time text summarization system to process conversational datasets.
- **Approach**: Fine-tuned Google Pegasus on the SAMSum dataset, designed data ingestion and transformation pipelines with 99.8% quality compliance, and deployed using FastAPI + AWS CI/CD for scalability.
- **Impact**: Boosted ROUGE-L by 18% and delivered real-time summarization with 200ms average latency handling 150 requests/sec.

Fake News Detection:

- **Objective**: Detect and classify fake vs. real news articles using machine learning.
- **Approach**: Preprocessed data (imputation, tokenization, lemmatization), engineered NLP features with TF-IDF & CountVectorizer, and trained multiple classifiers including Logistic Regression, Random Forest, and Decision Tree.
- **Impact**: Achieved 97% accuracy and 0.95 F1-score with Logistic Regression, reducing feature sparsity by 30% and improving detection reliability.

Awards

- Ehsaas Undergraduate Scholarship recipient
- Organized Mind Freezer Quiz Competition (first edition, record participation)

Involvement

Management Head – Calligraphy & Fine Arts Society:

- Coordinated **teams and activities**, ensuring smooth execution of events.
- Managed **budgets, sponsorships, and marketing** to support society growth.
- Organized **workshops and exhibitions**, boosting engagement and showcasing talent.

Event Head – COMSATS Science Club:

- Organized the **Mind Freezer Quiz Competition** (first edition, record participation).
- Developed **verbal and non-verbal quiz content** with the team.
- Managed **event logistics** including registration, scheduling, and on-site coordination.