

BILAL MOHAMMAD AFZAL

AI Engineer | Generative AI | RAG | AI Agents | Full-Stack AI Systems

bilal.mafzal7894@gmail.com | [Github](#) | [LinkedIn](#) | +92 317 4940014 | Lahore, PK

PROFESSIONAL SUMMARY

AI Engineer with 2 years of experience building Generative AI, RAG, AI agent, and machine learning applications. Hands-on experience with LLM orchestration, retrieval pipelines, FastAPI backend services, vector databases, and full-stack AI development. Strong in Python, LangChain, LangGraph, and deploying end-to-end AI solutions using modern backend, frontend, and MLOps tools.

WORK EXPERIENCE

LIZTEK — AI/ML Engineer [Full Time - Contract] | Lahore, Pakistan | May 2026 – Sep 2026

- **Built** end-to-end LLM agent, RAG, and **AI automation workflows** using LangChain, LangGraph, embeddings, and vector databases, **reducing** manual processing time by ~35% across supported workflows.
- **Developed** Python/FastAPI backend services and REST APIs with **PostgreSQL** and Redis, improving AI **workflow** response time by ~25% through optimized data access and caching.
- Implemented **Docker**, CI/CD, **automated** testing, and model evaluation workflows, reducing deployment effort by ~40% and **improving** consistency across **development** and release **cycles**.

DEVELOPERSHUB CO — AI/ML Engineering Intern [Full Time] | Mar 2026 – May 2026

- **Developed** and evaluated classification and **regression** models using Python and Scikit-learn, achieving up to **82%** accuracy across assigned machine learning use **cases**.
- Built and refined **LLM-based** chatbot workflows using **prompt engineering** and structured response techniques, improving **response** relevance by approximately 30% during evaluation.
- **Performed** feature **engineering**, data preprocessing, hyperparameter tuning, and model **optimization**, improving model **performance** by **8–15%** across experimentation cycles.

CODEALPHA — Junior ML Engineer [Contract] | Oct 2025 – Feb 2026

- Built a **churn prediction** pipeline using customer and **booking data**, enabling proactive retention actions and improving **customer retention** by 10%.
- **Engineered** Python ETL jobs on a **Linux VPS** to **ingest multi-source API data**, calculate operational and geolocation KPIs, and **load analytics-ready datasets**, reducing manual data preparation by **80%**.
- Built **automated Power BI dashboards** for Sales, HR, revenue, and operations, cutting recurring reporting effort by **60%** and giving 4 business functions centralized visibility into key KPIs.

PROJECTS

Financial Filings RAG Copilot — FastAPI | LangChain | Gemini | Qdrant | MongoDB | Streamlit | Docker

- Built an end-to-end **RAG pipeline** for SEC 10-K/10-Q filings covering parsing, **section-aware chunking**, **embeddings**, vector retrieval, **citation-grounded generation**, and FastAPI services.
- Experimented with **chunking**, metadata filtering, top-k retrieval, **refusal handling**, and **evaluation workflows** to improve retrieval quality and reduce unsupported answers.

AI Voice Calling Agent — LangGraph | Whisper | DeepSeek | Pocket TTS | FastAPI | MongoDB

- Built a prototype outbound voice agent using **LangGraph** workflows, **VAD**, **Whisper STT**, LLM orchestration, **TTS**, MongoDB prospect context, and **FastAPI/Zoom call control**.
- Tested conversation flows, **cancellation**, playback accounting, **MongoDB retries**, dry-runs, and **production-context rehearsals**, with GPU/audio latency and supervised outbound acceptance tracked as release gates.

FinFlow AI (In Progress) — Next.js | FastAPI | PostgreSQL | Redis | LangGraph | Qdrant | Docker

- Designing an **AI-powered financial operations platform** for SMBs with agent workflows for invoice and receipt processing, transaction categorization, bank reconciliation, and **anomaly detection**.
- Architecting enterprise features including **multi-tenancy**, **RBAC**, **audit logs**, with **CI/CD** and **Docker** deployment.

SKILLS

- **Generative AI & Agents:** RAG, AI Agents, LangChain, LangGraph, Prompt Engineering, Tool Calling, Embeddings, Grounded Generation
- **ML, Retrieval & Voice:** PyTorch, Scikit-learn, TensorFlow/Keras, NLP, Computer Vision, Qdrant, Pinecone, FAISS, Chroma, Whisper, VAD, STT/TTS
- **Backend, Data & Delivery:** Python, SQL, FastAPI, REST APIs, PostgreSQL, MongoDB, ETL, n8n, Make.com, Docker, Git/GitHub, Linux/VPS, Testing, Streamlit, Power BI.

PROFESSIONAL DEVELOPMENT & CERTIFICATIONS

- **Certifications:** Google Data Analytics Professional Certificate — Coursera | Machine Learning Specialization — DeepLearning.AI
- **Professional Development:** Apache Superset contributor (PR #5025) | SOFTEC 2026 | Computing Fest 2026 (UHE) | BNU ML Competition

EDUCATION

Virtual University | Bachelor of Science in Computer Science