# Mohammad Afzal

Machine Learning Engineer | Cloud Solutions Architect | AI-Driven Product Developer [mohammadafzal.tech@gmail.com](mailto:mohammadafzal.tech@gmail.com) | +917222876247 | [https://afzalex.com](https://afzalex.com/)

## Summary

Experienced **Machine Learning Engineer** with 10+ years of success in designing and deploying AI-driven **python** based systems at scale. Specialized in **LLM orchestration** using **LangChain** and **LangGraph**, building **RAG pipelines** with FAISS, and deploying production-grade applications on **AWS**, **Docker**, and **Kubernetes**. Strong command of **MLOps workflows**, including model lifecycle management with **MLflow**, pipeline orchestration with **Kedro**, and model serving via **SageMaker**. Proven ability to lead cross-functional engineering teams, optimize cloud infrastructure, and deliver intelligent solutions that drive measurable business impact.

## Technical Proficiencies

**Programming Languages**: Python, Java, JavaScript, SQL, Bash

**Machine Learning & MLOps**: LangChain, LangGraph, MLflow, Kedro, Scikit-learn, PyTorch (basic), JupyterLab

**Data Engineering & Processing**: Pandas, NumPy, Apache Kafka, Airflow, Snowflake, PostgreSQL, MongoDB, Redis, S3

**Backend Development**: FastAPI, Flask, Spring Boot, Node.js, SQLAlchemy, REST APIs

**Frontend Technologies**: ReactJS, AngularJS, Bootstrap

**Cloud Platforms**: AWS (EC2, Lambda, EKS, RDS, SageMaker, S3, CloudWatch), Databricks, Google Colabs

**DevOps & Deployment**: Docker, Kubernetes, Jenkins, GitHub Actions, Terraform (intro), CI/CD Pipelines

**Testing & Automation**: PyTest, JUnit, Selenium, BDD (Behave, Cucumber)

## Experience

### Lead Cloud Engineer, Technumen | 2024 - Present

* Led the development of a **voice-activated One Rule assistant**, combining **Python, Whisper, LangChain, LangGraph** and **ElevenLabs** for real-time conversational AI - [VoiceChain Template](https://github.com/afzalex/voicechain-template" \t "_new).
* Designed **multi-modal agent workflows** with LangGraph to handle FAQs, scheduling, and document summarization with stateful intent switching and fallback logic.
* Built **RAG pipelines** with **FAISS** for vector-based retrieval from enterprise knowledge bases.
* Deployed services using **Docker, AWS Lambda** and **CloudFront** to ensure scalability and low latency.
* Created an **LLM-powered resume screening system** in Python, integrating LangChain and LangGraph to evaluate resumes via multi-agent orchestration against job descriptions.

### Product Development Engineer, Calibo | 2020 - 2023

* Designed and deployed **Python-based microservices** using **FastAPI** and **SQLAlchemy**, focused on ML lifecycle management, real-time analytics, and scalable inference delivery.
* Built **streaming data pipelines** with **Apache Kafka** and **Spark Structured Streaming** for low-latency event ingestion and feature engineering at scale.
* Implemented modular ML workflows using **Kedro**, enabling reproducibility and collaboration across cross-functional teams.
* Managed end-to-end **MLOps pipelines** with **MLflow** for tracking experiments, managing model versions and deploying to **AWS SageMaker** for training and batch/real-time inference.
* Implemented automated **model retraining pipelines** based on data drift detection and integrated **ELK Stack dashboards** for performance monitoring and operational insights.
* Utilized **Snowflake** for scalable analytics and feature storage, supporting large-scale ETL and efficient real-time data retrieval.
* Led deployment automation using **Docker**, **Kubernetes**, and **Jenkins**, managing CI/CD pipelines and multi-environment provisioning.

### Senior Associate, Nagarro | 2019 - 2020

* data not available

### Senior Software Engineer, Enbraun | 2015 - 2019

* data not available

### Freelancer | Before 2015

data not available

## Project Profile

### OneRule – Liberty Mutual

An enterprise-grade, real-time voice assistant platform designed to enable natural, speech-driven interactions for internal knowledge access, scheduling, and operational support. Built with a multi-modal architecture combining automatic speech recognition (ASR), large language models (LLMs), and text-to-speech (TTS) synthesis.

Key Contributions:

* Architected the assistant’s core pipeline in **Python**, utilizing **Whisper (ASR)**, **LangChain + LangGraph (agent orchestration)**, and **ElevenLabs (TTS)** to deliver responsive and natural voice-based interactions.
* Integrated **Retrieval-Augmented Generation (RAG)** with **FAISS** for grounding LLM responses in private enterprise knowledge bases, ensuring accuracy and compliance.
* Deployed the solution using **Docker**, **AWS Lambda**, **API Gateway**, and **CloudFront** in a fully serverless, horizontally scalable environment.
* The system evolved from an open-source prototype, **VoiceChain Template (GitHub)**, into a production-ready assistant serving enterprise users : [VoiceChain Template – GitHub](https://github.com/afzalex/voicechain-template" \t "_new)

### LAZSA – Product Platform-as-a-Service

Industry-first Product Platform-as-a-Service (pPaaS) to redefine digital innovation and drive competitive advantage through digital business. The platform provides a comprehensive set of tools and technology stack to efficiently manage, define, design, develop, and deploy phases of product development.

Key Contributions:

* Led architecture and development of core platform modules using **Java 8 (Spring Boot)** and **ReactJS, Python**, ensuring modular design and extensibility across services.
* Developed ML and data pipeline components in **Python**, integrating **Kedro** for pipeline management and **MLflow** for experiment tracking and reproducibility.
* Integrated **AWS Services** (S3, Lambda, SageMaker) and **Kafka** to support event-driven workflows and real-time communication between loosely coupled services.
* Orchestrated platform deployment using **Docker**, **Kubernetes**, and **Jenkins**, building robust CI/CD workflows for efficient release cycles.
* Built observability dashboards using the **ELK Stack** and collaborated with stakeholders to ensure scalability, compliance, and performance optimization across the platform.

### Passbook – Northern Trust

Enterprise-grade modernization initiative aimed at decoupling legacy systems into secure, maintainable microservices while integrating robust DevOps and identity management practices across the financial services infrastructure.

**Key Contributions**:

* Led the development and optimization of core services using **Spring Boot** and **FastAPI**, ensuring modularity, fault isolation, and consistent API governance.
* Implemented centralized service discovery and configuration using **Eureka Server** and **Spring Cloud Config**, improving system visibility and deployment agility.
* Integrated **Azure Active Directory (AD)** for secure, federated authentication and enforced SSO across internal applications.
* Built CI/CD pipelines using **GitHub Actions** and **Jenkins**, enabling automated builds, test execution, and production-grade deployments.
* Performed performance benchmarking using **JMeter**, and wrote behavior-driven test suites in **Behave** and **PyTest**, ensuring system reliability and adherence to business logic.

### DiPEC (Digitalized Project Execution)

data not available

### eResource Scheduler

data not available

### Abho Costing Sheet

data not available

## Education Detail

data not available