# 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 9+ years of success in designing and deploying AI-driven 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, AWS SageMaker, Kedro, Scikit-learn, TensorFlow, Hugging Face Transformers, PyTorch (basic), JupyterLab, XGBoost, LightGBM

**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, 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)

**Tools & Utilities**: VS Code, Postman, Git, Linux CLI, Streamlit, ELK Stack, SQLite

## Experience

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

* Spearheaded development of a *voice-activated* ***One Rule*** in production, building on concepts prototyped in the open-source [VoiceChain Template](https://github.com/afzalex/voicechain-template" \t "_new).
* Engineered a robust multi-modal system combining Whisper for speech-to-text, **LangChain + LangGraph** for contextual agent reasoning, and ElevenLabs for speech synthesis, enabling dynamic, real-time conversations.
* Designed agent workflows using LangGraph to support stateful interactions, intent switching, and fallback handling across tasks like FAQ resolution, scheduling, and document summarization.
* Integrated RAG-based pipelines with vector search (FAISS) to answer user queries grounded in enterprise knowledge bases.
* Deployed the solution using Docker, AWS Lambda, and CloudFront to ensure scalability, low latency, and secure API exposure.
* Collaborated with product teams to refine UX and voice flow, optimizing turn-taking and naturalness of responses.
* Developed an LLM-powered resume screening system utilizing **LangGraph** **+ LangChain**, enhancing the recruitment process by dynamically retrieving and processing resumes using RAG.
* Implemented a multi-agent orchestration framework to evaluate resumes against job descriptions, providing a rating system and feedback mechanism.

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

* Designed and deployed scalable, data-driven microservices using Python(FastAPI) and SQLAlchemy, focusing on ML model lifecycle, real-time analytics, and inference delivery.
* Engineered **streaming data pipelines** using **Apache Kafka** and **Spark Structured Streaming** to support low-latency event ingestion, feature generation, and model serving 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.
* Developed automated model retraining jobs triggered by data drift signals, and integrated performance monitoring dashboards using ELK Stack.
* Leveraged **Snowflake** for scalable analytics and feature storage, ensuring efficient retrieval and transformation of large datasets in production environments.
* Led deployment automation using **Docker**, **Kubernetes**, and **Jenkins**, orchestrating CI/CD pipelines and environment provisioning across staging and production clusters.

### Senior Associate, Nagarro | 2019 - 2020

* Led the digital transformation of project development by developing web applications using FastAPI, RabbitMQ, Angular 8, and Postgresql.
* Built microservices using **Spring Boot**, **RabbitMQ**, and **Angular 8** for scalable, event-driven applications.
* Optimized PostgreSQL queries and backend performance for real-time dashboards.
* Contributed to Agile delivery, TDD with JUnit, and API integration across distributed systems.
* Gained foundational experience in backend systems and asynchronous workflows, later extended into ML and data engineering roles.

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

* Developed multi-user resource management and scheduling software using JavaEE, AngularJS, and Restful services.
* Designed and implemented RESTful APIs for seamless integration with external systems.
* Collaborated with clients to gather requirements, provide technical guidance, and deliver custom solutions.
* Led the migration of legacy applications to modern technologies, resulting in improved performance and user experience.

### Freelancer | Before 2015

* Successfully completed various freelance projects, including web application development and customization based on client requirements.
* Collaborated closely with clients to understand their needs and deliver tailored solutions within agreed timelines.

## 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 using **Whisper** (ASR), **LangChain + LangGraph** (agent orchestration), and **ElevenLabs** (TTS), ensuring seamless, low-latency interactions.
* Designed modular, stateful agent workflows in **LangGraph**, allowing for complex task switching, context retention, and recovery in long-running conversations.
* Integrated **Retrieval-Augmented Generation (RAG)** with **FAISS** to ground responses in real-time against private knowledge bases, enhancing factual reliability and compliance.
* Deployed the solution in a fully serverless, production-grade environment using **Docker**, **AWS Lambda**, **API Gateway**, and **CloudFront**, ensuring horizontal scalability and global availability.
* Tuned the voice flow and user experience in collaboration with UX teams, resulting in a highly responsive, human-like interface for enterprise users.
* The system was built upon core concepts initially prototyped in the open-source project: [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** ensuring high cohesion and extensibility across services.
* Integrated **AWS Services** (S3, Lambda, SageMaker) and **Kafka** to support event-driven workflows and real-time communication between loosely coupled services.
* Implemented modular ML pipelines using **Kedro**, tracked experiments and deployments with **MLflow**, and used **SageMaker** for model training and inference workflows.
* Deployed the platform using **Docker** and **Kubernetes**, with CI/CD pipelines orchestrated through **Jenkins**, enabling seamless rollouts and rollbacks.
* Developed telemetry and observability dashboards using **ELK Stack**, ensuring operational transparency and system health monitoring.
* Collaborated with stakeholders to ensure scalability, compliance, and performance optimization, enabling rapid feature delivery and consistent product quality.

### 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.
* Coordinated with cross-functional teams to align platform capabilities with enterprise compliance standards and SLAs.

### DiPEC (Digitalized Project Execution)

Web application created to digitalize project development, utilized by engineers of Siemens. The project involved team leadership and development from the initial stages. The application was developed in a microservices architecture, ensuring scalability.

Key Contributions:

* Led a team in the development of features and functionality, meeting project deadlines and ensuring high code quality.
* Leveraged Spring Boot, Kafka, Angular 8, Postgresql, and Material UI to deliver a robust and user-friendly application.

### eResource Scheduler

Enbraun's eResource Scheduler is a multi-user feature-rich software for resource management and scheduling. The application enables organizations to efficiently schedule, plan, and manage employees, equipment, and other resources.

Key Contributions:

* Developed RESTful services, utilizing AngularJS, Bootstrap, OAuth2, and JavaEE to create a comprehensive resource management solution.
* Collaborated with clients to gather requirements, design customizations, and provide technical guidance.
* Implemented user-defined fields, booking charts, time sheets, and other advanced features to enhance resource management capabilities.

### Abho Costing Sheet

Web application designed to manage costing in the fashion designing industry. The application enables management of clients, fabrics, labor costs, dyeing, overheads, and provides precise cost analysis.

Key Contributions:

* Developed RESTful services using AngularJS, Bootstrap, Servlets, Java EE, and Wildfly to create a user-friendly and efficient costing management application.
* Implemented role-based access control to ensure appropriate authorization for different application components.

## Education Detail

**Master of Science** in Data Science from Liverpool John Moores University (Distance)

**Post Graduate Diploma** in Data Science from IIITB Bangalore (Distance)

**Bachelor of Technology** in Computer Science from Uttarakhand Technical University

**Higher Secondary School** from Modern School Rishikesh (CISCE)

**High School** from Modern School Rishikesh (CISCE)