Mohammad Afzal

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

Summary

Machine Learning Engineer with 10+ years of experience building and deploying **Python-based AI systems and microservices** at scale. Skilled in **LLM orchestration** (LangChain, LangGraph), **RAG pipelines** (FAISS), and cloud-native deployments on **AWS, Docker, and Kubernetes**. Strong in **MLOps** using MLflow, Kedro, and SageMaker. Proven leader in delivering intelligent, scalable solutions that drive real 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

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.
- 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.
- 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, Kafka, Databricks** 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

- Developed enterprise applications using a **hybrid microservices stack—Node.js** for lightweight APIs and **Kotlin and Java with Spring Boot** for core business logic.

- Led a small team of engineers in designing, developing, and maintaining microservices, ensuring high scalability and modularity.
- Led the digital transformation of project development by developing web applications using **Kotlin** + **Spring Boot**, **RabbitMQ**, **Angular 8**, and **Postgresql**.
- Actively participated in agile development processes, including sprint planning, daily stand-ups, and retrospective meetings.

Senior Software Engineer, Enbraun | 2015 - 2019

- **Core contributor and technical lead** on the development of *eResource Scheduler*, a multi-user resource management platform built from the ground up for scalability, availability, and customization on **AWS (EC2, RDS, S3, ALB)**.
- Designed and developed **RESTful services** and dynamic frontend modules using **JavaEE**, **AngularJS**, and **Bootstrap**, forming the backbone of the platform's booking, timesheet, and reporting features.
- Mentored junior developers and facilitated code reviews, fostering a culture of quality and hands-on learning within the team.
- Played a core role in the platform's **first public release after 3 years of active development**, contributing across all phases—from designing architecture and building features to testing, optimization, and deployment of a **highly available and scalable system**.

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

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)
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.
Link : https://www.dipec.siemens.co.in
Key Contributions:

- **Led the end-to-end development** of critical features from inception, working closely with Siemens stakeholders to translate engineering workflows into intuitive application modules.
- Designed and implemented services using **Spring Boot** and **Kafka** to enable scalable, event-driven communication between modules.
- Developed a dynamic and responsive frontend using **Angular 8** and **Material UI**, integrated seamlessly with **PostgreSQL** for reliable data persistence.
- Took ownership of code quality, reviews, and sprint delivery, ensuring adherence to agile processes and successful on-time releases.

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.
Link : https://www.eresourcescheduler.com
Key Contributions:

- Developed scalable **RESTful APIs** and interactive frontend modules using **JavaEE**, **AngularJS**, **Bootstrap**, and **OAuth2**, enabling secure, responsive, and modular scheduling capabilities.
- Implemented advanced features including **user-defined fields**, **dynamic booking charts**, **timesheet modules**, and **customized reporting**, significantly improving planning flexibility and user experience.

ExamChi – Online Exam Practice Platform
A cloud-based exam preparation tool designed to deliver dynamic quizzes, performance tracking, and practice sessions for students and professionals. Built as a full-stack application to demonstrate scalable architecture, real-time user interaction, and cloud-native deployment.
Link : https://examchi.afzalex.com/
Key Contributions:

- Designed and developed the platform using **Node.js** (Express) for backend services, **ReactJS** for the frontend, and **MongoDB** for schema-flexible data storage.
- Deployed on **AWS** with a focus on performance and availability, leveraging services like EC2 and S3 for hosting and storage.

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