

Mohammad Safwan

Dubai, United Arab Emirates | +971 50 323 4783

msafwan2002@gmail.com | [LinkedIn](#)

PROFESSIONAL SUMMARY

AI/ML Engineer with robust experience building and deploying machine learning-powered applications and LLM-integrated platforms. Proven track record improving AI accuracy by 45%, system performance by 30%, and backend scalability through advanced RAG architectures, vector databases, and cloud-native tools. Strong foundation in Python, NLP, model fine-tuning, and scalable API development.

EDUCATION

MSc in Artificial Intelligence

Heriot-Watt University

Dubai, UAE

Jan 2026 – Present

B.Tech in Computer Science

SCMS School of Engineering and Technology

Kerala, India

May 2020 – May 2024

TECHNICAL SKILLS

Programming Languages: Python, JavaScript, SQL, C#

Frameworks & Libraries: FastAPI, Flask, React, TensorFlow, Pandas

AI/ML Technologies: Large Language Models (LLMs), OpenAI API, LangChain, LlamaIndex, Retrieval-Augmented Generation (RAG), Natural Language Processing (NLP), Prompt Engineering, Machine Learning, Deep Learning

Data Engineering: Apache Kafka, Fluentd, Data Pipelines, ETL

Databases: PostgreSQL, SQLite, Vector Databases

DevOps & Tools: Docker, Git, GitHub, Postman, CI/CD, REST APIs

Cloud Platforms: Amazon Web Services (AWS), Microsoft Azure

PROFESSIONAL EXPERIENCE

Associate Software Developer

Jul 2024 – Nov 2024

Quantum Acumen Pvt. Ltd.

Remote, India

- Engineered and deployed LLM-based analytics system enabling natural language querying of structured databases, streamlining data access for non-technical users
- Designed retrieval-augmented generation (RAG) pipelines with vector search integrations, improving LLM response accuracy by 45% through optimized context retrieval
- Built real-time backend systems and RESTful APIs using FastAPI and Python, achieving 30% reduction in system runtime through performance optimization
- Collaborated with cross-functional teams to refine prompt engineering strategies and AI feature design for enhanced user experience

Software Developer Intern

Jan 2024 – Jul 2024

Quantum Acumen Pvt. Ltd.

Remote, India

- Optimized database schemas and backend logic using Python and FastAPI, significantly enhancing API performance and response times
- Supported AI model fine-tuning and prompt optimization efforts across multiple NLP tasks, improving model output quality
- Contributed to data pipeline development using Kafka for real-time data streaming and Fluentd for centralized logging infrastructure

PROJECTS

Sentiment Analysis Using NLP

- Built text classification model for sentiment detection (positive/negative) using natural language processing techniques
- Applied tokenization, word embeddings, and traditional ML algorithms achieving effective classification performance

Plant Disease Detection App

- Developed Convolutional Neural Network (CNN)-based model to identify plant diseases from leaf images with high accuracy
- Deployed solution as mobile-accessible application, enabling farmers to diagnose plant health issues in real-time