SIYAM FAHAD

TECHNICAL LEAD | AI & SOFTWARE SYSTEMS ARCHITECT

Dhaka, Bangladesh | +880 1643 330094 | Siyamweb3@gmail.com |

PROFESSIONAL SUMMARY

Technical leader and Microsoft Certified AI Engineer with 7+ years of experience architecting and delivering production-grade AI, SaaS, and embedded systems for organizations across Germany, the UK, and MENA. Hands-on expertise in Python, scalable service architecture, Docker, Kubernetes, AWS SageMaker, RAG, LLM applications, NLP, and production deployment. Current CTO with experience setting technical direction, translating business needs into software solutions, and building systems spanning enterprise automation, payments, voice AI, and generative media.

CORE COMPETENCIES

Leadership & Architecture: Technical strategy, software architecture, AI SaaS platforms, production systems, solution design
Engineering: Python, Docker, Kubernetes, API-integrated services, cloud deployment, embedded Linux/Raspberry Pi
AI/ML: Generative AI, RAG, LangGraph, LangChain, LlamaIndex, PyTorch, TensorFlow, Hugging Face, NLP
Infrastructure & Data: AWS SageMaker, Pinecone, Chroma, FAISS, scalable retrieval pipelines, model optimization

PROFESSIONAL EXPERIENCE

Chief Technology Officer | AURAS Technologies MENA FZCO | 2023–Present

- Directed end-to-end technical strategy and architecture for AI-driven SaaS platforms that automated business workflows using OpenAI GPT models and custom open-source LLMs.
- Designed a production RAG knowledge-retrieval system for enterprise use, improving the relevance and accuracy of information delivered to client workflows.
- Engineered a blockchain-integrated payment gateway and non-custodial cryptocurrency wallet using Web3 technologies and smart contracts.
- Translated product and business objectives into technical roadmaps spanning application architecture, AI integration, deployment, and platform evolution.

AI Systems Engineer (Part-Time) | Enny-Thing Ltd, UK | 2022–2025

- Built the S61 Voice Assistant, an automotive AI system combining real-time diagnostics, voice recognition, and predictive-maintenance capabilities.
- Optimized and deployed NLP pipelines and Transformer models on resource-constrained Raspberry Pi hardware, balancing latency, reliability, and compute limitations.
- Troubleshot integration and performance constraints across embedded hardware, speech-processing components, and AI inference workflows.

Senior Generative AI Engineer | Youvendo GmbH, Germany | 2021–2023

- Led development of an automated article-generation engine using GPT and BERT, increasing content-production efficiency by 40%.
- Designed and deployed Mokki Art Generation using Stable Diffusion to transform facial images into high-quality personalized digital artwork at large scale.
- Delivered production generative-AI capabilities for customer-facing products, covering solution design, model integration, optimization, and deployment.

Machine Learning Engineer | Skylar Softwares | 2020–2021

- Developed a Danish automatic speech-recognition system using DeepSpeech for a low-resource language dataset.
- Researched and implemented deep-learning techniques to improve voice-assistant latency and user experience.

SELECTED TECHNICAL CAPABILITIES

Languages & Development: Python; AI/ML application development; service and workflow integration
Cloud & Containers: Docker, Kubernetes, AWS SageMaker, production deployment
Generative AI: GPT-4o, Claude 3.5, Llama 3, RAG, LangGraph, LangChain, LlamaIndex, LoRA, QLoRA
ML Frameworks: PyTorch, TensorFlow, Transformers, Hugging Face, BERT, DeepSpeech
Data & Retrieval: Pinecone, Chroma, FAISS, vector search and retrieval optimization
Specialized Systems: Speech-to-text, Whisper, TTS, computer vision, Stable Diffusion, ControlNet, embedded AI

CERTIFICATIONS & PROFESSIONAL DEVELOPMENT

- AI & ML Engineering Professional Certificate — Microsoft
- Generative AI for Software Development — DeepLearning.AI
- Python Specialization — University of Michigan
- Complete Generative AI Course with LangChain and Hugging Face — iNeuron.ai
- PyTorch for Deep Learning Bootcamp — Zero To Mastery Academy