

Qossay Abuamriya

qossay.amriya@gmail.com | +971551538289 | Dubai, UAE | [LinkedIn](#) | [GitHub](#)

Profile Summary

Software Engineer with 3 years of experience, specializing in AI and cloud driven solutions that address business and user challenges, with expertise in leveraging AI services from major providers such as Azure, OpenAI, IBM, AWS, and Google. Experienced in NLP, OCR, STT, TTS, and LLMs, and adept at developing software for cloud environments to maximize AI's impact.

Professional Experience

Business Digital Transformation Engineer

Takaful Emarat Insurance PSC.

Jun. 2024 – Current

On-sight, Dubai

- Developing an **advanced Identity Document Processing solution** that handles **mixed document** IDs uploaded by end users in a **single file**. The solution efficiently processes multiple instances of the same document for one or more individuals and handles different types of identity documents with varying alignments, resolutions, and file types. Leveraging Azure AI Services, such as **Azure OCR**, **Azure Document Intelligence**, and **Azure OpenAI LLMs**, a seamless **pipeline** was created to accurately extract specific details for a certain user, from the documents. The solution emphasizes **cost-effectiveness** by refining the data passed to LLMs and selecting the **optimal models**, ensuring a reliable and efficient document processing workflow. **This results in saving about 50% of the documents' processing time.**
- Developing an **AI-powered customer support chatbot** solution using **LLMs** to understand and recognize customer intent and requests accurately. The chatbot generates **tailored, well-structured responses** based on individual user data while supporting various actions through seamless integration with primary user portals. Utilizing **generative AI** and LLMs from **OpenAI**, the solution also incorporates advanced control features like **Azure Content Filtering** and responsible AI practices, ensuring **safe and reliable** interactions, supported by **Azure's Responsible AI tools**. **Reducing the TAT by 70%.**
- Utilizing Azure App services and container services to deploy and configure applications on **Azure web apps** and **Container Apps**. Configuring **CI/CD pipelines** and setting up ingress controllers and other configurations, **reducing time-to-market by 25%.**

Software Engineer

CoolRiots - IBM Partner.

Oct. 2022 – Current

Remote, Singapore

- Participated in developing an **Agentic AI-powered Business Execution Platform** that leverages **generative AI** to address diverse business challenges across industries such as hospitality, law firms, and education centers.
- Applied **fine-tuning** techniques on IBM WatsonX Studio, Azure OpenAI, and OpenAI to **ground Large Language Models (LLMs)** in specific fields and topics, resulting in a **40% improvement in model accuracy.**
- Utilized **Retrieval-Augmented Generation (RAG)** techniques with a **vector database** and the **Langchain Python library** to customize LLM responses based on specific datasets and control the **contextual relevance of the model's outputs.**
- Leveraged **Groq APIs** to optimize the **inference speed of LLMs**, improving the performance of applications by **30%.**
- Implemented **Kubernetes (AKS, IKS, and OpenShift)** for orchestrating various application deployments, ensuring a remarkable **99.95% uptime and availability.** Demonstrated expertise in configuring **NGINX** for high-availability setups, adhering to **both IBM Cloud and Azure's operational standards** and best practices.

- Leveraged **Terraform** to automate the provisioning and management of **cloud infrastructure and Kubernetes resources**, enhancing deployment efficiency and consistency across **multiple environments**. This strategic implementation streamlined operations and reduced setup times by over 30%.
- **Utilized IBM Cloud Object Storage (COS) and IBM WatsonX** to design and implement scalable storage and AI-driven solutions, optimizing data accessibility and enhancing AI capabilities. These solutions streamlined data processing and reduced operational overhead, improving overall system efficiency by 15%.

Skills

Technical skills: LLMs inference, Prompt Engineering, Fine Tuning, RAG, Azure PaaS, Azure SaaS, IBM Cloud, REST APIs, Git, Bash, Python, FastAPI, ASP.Net, SQL, Docker, Scrum, Performance and Cost Optimization, Kubernetes, Red Hat OpenShift, GenAI, OpenAI, Gemini, Claude sonnet 3.5, cursor AI, AI agents.

Soft Skills: Project management, Leadership, Adaptability, Problem-Solving, Public speaking, Agile teamwork, Customer-centric attitude, Passion for automation.

Languages: Arabic (Native), English (professional).

Awards and Certificates

Microsoft Certified: Azure AI Engineer Associate	Issued by Microsoft· Jan. 2025
Microsoft Certified: Azure Administrator	By Microsoft. Under-going
Certified Kubernetes Administrator	By Linux Foundation. Under-going
Certified Kubernetes Administrator	Issued by Udemy. Feb. 2024
Microsoft Certified: Azure Artificial Intelligence Fundamentals	Issued by Microsoft· May. 2023
Microsoft Certified: Azure Security, Compliance, and Identity	Issued by Microsoft· Mar. 2023
Microsoft Certified: Azure Data Fundamentals	Issued by Microsoft· Feb. 2023
IBM Badge: Watson Assistant Sales Foundation	Issued by IBM· May. 2023
IBM Badge: Watson Assistant Technical Sales Intermediate	Issued by IBM· May. 2023

Education

University Malaysia of Computer Science & Engineering (UNIMY)	<i>Kuala Lumpur, Malaysia</i>
Bachelor’s degree in software engineering (Hons) 3.73 CGPA	