

Ala'a Alawdi

AI Agents Systems Engineer | LLM-Driven Autonomous Solutions Architect

Yemen, IBB, Thirty Street | +967776315292 | alawdisoft@gmail.com | [Linkedin](#) | [Kaggle](#) | [Github](#)

Bio

I am an AI and automation engineer skilled in building intelligent agents and workflow systems. I've developed multi-agent automation using tools like n8n, Zapier, and NTL for tasks like email replies and marketing. My work includes advanced desktop automation, WhatsApp bots, and AI-powered grocery managers. I created translation agents for Arabic-English academic texts, preserving structure and terminology. I've optimized RAG systems and developed AI chatbots with file extraction and vector search. My experience includes fine-tuning LLMs with LoRA, quantization, and knowledge distillation for Arabic NLP. I also built a medical diagnosis tool using specialized agents and led AI projects in NLP, vision, and analytics.

Experience

SMA Soft | Chatbot Developer 2024/8-2025/2

Designed intelligent chatbots with context-aware, multi-turn conversations, integrating file text extraction and data retrieval features to enhance customer engagement and automate workflows.

Freelance AI Developer 2022-2024

Worked as a freelance AI Developer, specializing in language model fine-tuning, deep learning for NLP and image classification, and data analytics using SQL and Power BI.

Education

IBB University, BS Information Technology 2020/2024

Major: Science | **Minor:** Computer Science

Grade: Very Good **Project:** Excellent.

Skills & abilities

Programming Languages: Python, SQL and JavaScript.

Databases: SQL (PostgreSQL, MySQL) and NoSQL (MongoDB).

Code Quality: Code reviews, SOLID principles, design patterns.

System Design: Scalability, load balancing, caching (Redis), database sharding.

API Testing & Documentation: Postman, Dogapi.

Model Fine-Tuning & Optimization: LoRA, LLaMA-Factory, vLLM, Unsloth.

AI & Chatbot Development: LangGraph and Retrieval-Augmented Generation (RAG).

Mathematics & Statistical Analysis: Linear Algebra, Calculus, Probability, Statistics.

Web Frameworks: FastAPI, Django, Streamlit and Docker.

Soft Skills: Problem-Solving, Critical Thinking, Communication, Continuous Learning, Attention to Detail, Creative & Analytical Thinking

LANGUAGES

English Highly proficient

Arabic Native speak

Projects

AI-Powered Workflow Agents for Automated Marketing and Communication

This project demonstrates the creation of intelligent agents using workflow automation platforms like n8n, Zapier, and NTL. The agents are designed to handle complete tasks such as automated email replies, marketing campaign management, and multitask execution. By integrating AI capabilities with low-code/no-code platforms, the system streamlines repetitive operations, enhances customer engagement, and improves overall productivity.

AI Multi-Agent Desktop Automation System

Engineered an AI agent-based system to automate desktop interactions using a collaborative framework of specialized agents: Orchestrator Agent, Vision-Language Agent (VLM Agent), Parser Agent (OmniParser), and Executor Agent.[link](#)

AI-Powered Book Translation Agent: Arabic-English Academic Content Transformer

Developed an advanced translation system that processes entire academic books from English to Arabic while preserving original formatting, document structure, and key technical terminology. This specialized translation agent outperforms standard translation tools by maintaining context across large documents and selectively preserving critical English terms within Arabic text for improved student comprehension.[link](#)

Grocery Management System Using AI Agents

Developed an AI-powered grocery management system using **CrewAI** to automate grocery tracking, expiration date estimation, and recipe recommendations. The system integrates multiple intelligent agents, each assigned a specific task to streamline household grocery management.[link](#)

Chatbot and Short Term Memory

AI Chatbot with OpenAI + LangGraph: Memory management via summarization, MVC-structured, Docker-containerized, production-ready with state debugging.[link](#)

WhatsApp Agent Automation Project

Developed a Python-based automation agent for WhatsApp using advanced tools and frameworks. Focused on creating a scalable and efficient solution for automating WhatsApp interactions.

Virtual Research Assistant

The Virtual Research Assistant is an AI-powered tool that helps researchers find, summarize, and analyze academic papers. It uses AI agents to automate summaries and highlight key pros and cons, aiding in informed decision-making.

Adaptive Local AI Platform for RAG Optimization

This fully local AI platform optimizes Retrieval-Augmented Generation (RAG), offering flexible AI model selection, embedding techniques, and text extraction while ensuring data privacy and security. It supports offline operations, integrates an Oracle database, and enables efficient, customizable AI-driven interactions.

AI Chatbot with File Text Extraction and Data Retrieval

Developed an AI-powered chatbot with integrated file text extraction using Streamlit, Llama.cpp, and ChromaDB, supporting multi-turn conversations and context-aware responses. It processes PDF, DOCX, CSV, and TXT files, embeds text into a vector database, and offers a seamless Streamlit interface with real-time updates.[link](#)

Efficient Fine-Tuning of DeepSeek-R1-Distill-Llama-8B for Arabic NLP Using LoRA, Quantization, and Knowledge Distillation

The project fine-tuned the DeepSeek-R1-Distill-Llama-8B model on Google Colab using a Tesla T4 GPU, Unsloth for efficient training, 4-bit quantization, and LoRA for parameter optimization. It follows Hugging Face standards, supports offline training on local hardware, and ensures data privacy.

Practical Fine-Tuning of Large Language Models for Arabic NLP with LoRA and vLLM

The project fine-tunes LLMs for Arabic NLP using LLaMA-Factory, Hugging Face Transformers, and vLLM, with LoRA adapters, 4-bit quantization, and Pydantic schemas for efficient training. It supports entity extraction, translation, and real-time local deployment, ensuring performance and cost-effectiveness through knowledge distillation and load testing.

Medical Diagnosis Application with Multidisciplinary Analysis

This Python-based application assists in medical diagnosis using specialized AI agents (Cardiologist, Psychologist, Pulmonologist) for a multidisciplinary analysis. It processes medical reports, executes agents concurrently, and delivers a comprehensive diagnosis via a Tkinter GUI.

Multi-Project AI Portfolio: Machine Learning, Deep Learning, NLP, Computer Vision, and Data Analytics

Developed and fine-tuned AI models in NLP, computer vision, and data analysis, utilizing AraBERT for Saudi dialects and deep learning models like MobileNet, ResNet50, and VGG for high-accuracy image classification. Applied LLMs for financial insights and used SQL and Power BI for advanced sales data analytics.

COURSES

Fine-Tuning Large Language Models (LLMs)

- Fine-Tuning LLMs Generative AI YouTube (freeCodeCamp)
- Generative AI Fine-Tuning LLM Models YouTube (Krish Naik)

Retrieval-Augmented Generation (RAG) Fundamentals and Advanced Techniques

- RAG Fundamentals and Advanced Techniques – Full Course YouTube (freeCodeCamp)
- Advanced RAG (Retrieval-Augmented Generation) Techniques YouTube

Machine Learning

- Machine Learning (StatQuest | Andrew Ng)
- Machine Learning (CM 763: Classification by Ali Ghodsi)
- Machine Intelligence (University of WorldQuant)

Deep Learning

Certifications

- Andrew Ng (YouTube)
- Neural Networks (MLScientist: Hesham Asem | YouTube)
- Articles (Medium, Towards Data Science, Machine Learning Mastery)

Mathematics

- Linear Algebra (Ahmed Fathy | Gilbert Strang - MIT)
- Calculus (AnaHr)
- Probability and Statistics (Khan Academy)

Data Analysis

- Python Data Analysis & Visualization (YouTube)

Other Resources

- Articles (Medium, Towards Data Science, Machine Learning)

Applied Data Science Lab (Jun 2023 – Jul 2023)

Completed eight end-to-end data science projects covering data extraction, cleaning, ETL pipeline creation, and building supervised/unsupervised ML models with data visualization.

Neural Networks and Deep Learning – DeepLearning.AI (Sep 2024 – Dec 2024)

Mastered deep learning fundamentals, including neural network architectures, backpropagation, and mathematical modeling for AI applications.

Developing Serverless AI Applications with Rust and AWS (Apr 2024)

Acquired skills in setting up a Rust development environment, utilizing the AWS SDK for Rust, and building AWS Lambda functions. Covered MLOps challenges, Generative AI workflows in Rust, and serverless architectures using AWS Bedrock and SageMaker. Practical labs included building microservices with Rust Axum and deploying serverless solutions on AWS.

Data Fundamentals – University of Michigan (Jan 2025)

Developed strong foundations in data analysis, covering key concepts in data structures, databases, and data processing techniques.

LLaMA for Python Programmers – University of Michigan (Jan 2025)

Learned efficient methods to fine-tune and deploy LLaMA models using Python, focusing on model optimization and integration with AI pipelines.

Developing Serverless AI Applications with Rust and AWS