

Ahmed Ibrahim Metawee Youssef

Mansoura, Dakahlia Governorate, Egypt

+201114119479 | aimt16@hotmail.com | <https://www.linkedin.com/in/aimty> | <https://sites.google.com/view/aimty>

Summary

Results-oriented Data Scientist and AI Consultant specializing in the development and deployment of scalable Generative AI solutions. Expertise in fine-tuning Large Language Models (LLMs) and architecting Retrieval-Augmented Generation (RAG) pipelines to enhance accuracy, reduce hallucinations, and deliver measurable business value. Proven ability to translate complex AI concepts into high-performing, production-ready systems for diverse client needs, bridging the gap between AI innovation and practical application.

Professional Experience

Freelance GenAI Consultant & Data Scientist | Self-Employed

2021 – Present

- * Designed, developed, and deployed end-to-end Generative AI solutions for various business use cases, focusing on transforming open-source models (e.g., LLaMA, Mistral, Falcon) into customized, high-performing assets.
- * Specialized in architecting and implementing Retrieval-Augmented Generation (RAG) systems using frameworks like LangChain and Haystack, integrated with vector databases (FAISS, Pinecone) to ensure grounded, factually accurate outputs.
- * Executed domain-specific fine-tuning strategies for LLMs, demonstrably reducing model hallucinations and improving response relevance for specialized tasks.
- * Achieved significant performance improvements in client projects, including enhancing retrieval precision by up to 40%.
- * Engineered scalable data and model deployment pipelines leveraging cloud platforms (AWS, GCP) and API delivery mechanisms, considering enterprise requirements for security and compliance.
- * Consulted directly with clients to understand requirements, define project scope, and architect tailored AI solutions aligning with strategic goals.
- * Optimized model inference times and resource utilization, contributing to lower operational costs for deployed AI systems.
- * Managed the full lifecycle of AI projects, from initial concept and feasibility assessment through to production deployment and ongoing optimization.

Skills

- * AI & Machine Learning: Generative AI, Large Language Models (LLMs), LLM Fine-Tuning (LLaMA, Mistral, Falcon), Retrieval-Augmented Generation (RAG), Natural Language Processing (NLP), Embedding Techniques, Model Optimization, Hallucination Reduction Strategies.

- * Frameworks & Tools: LangChain, Haystack, Python.

- * Data & Cloud Technologies: Data Pipeline Architecture, Vector Databases (FAISS, Pinecone), Cloud Platforms (AWS, GCP), API Development & Integration, Scalable Deployment (Cloud, On-Prem).

- * Consulting & Business: Solution Architecture, Client Needs Assessment, Project Scoping, Technical Strategy, Business Value Articulation, Problem-Solving, Independent Project Delivery.

Key Achievements

- * Improved retrieval precision in RAG systems by up to 40% through optimized architecture and fine-tuning.

- * Successfully transformed open-source LLMs into production-ready, domain-specific solutions for clients.

- * Designed and implemented scalable, end-to-end GenAI systems from concept to deployment.

- * Contributed to reduced operational costs for clients via optimized model inference.

Education

- * AI product manager, AI/ML, Bertlesmann, 2021

- * Data science learning path, LinkedIn learn, 2022