

Islam Abd-Elhady

📍 Asyut, Egypt ✉ eslamabdo71239@gmail.com ☎ +20 1066463355
in linkedin.com/in/islam-abd-elhady-323523211 🌐 github.com/Islam-hady9

Professional Summary

AI/ML Engineer with hands-on expertise in **Generative AI, LLM fine-tuning, agent-based systems, and computer vision**. Skilled in building **end-to-end pipelines** — from model training (LoRA, Diffusers, Transformers) to **scalable deployment** (FastAPI, Docker, Azure/AWS). Delivered production-ready solutions including multilingual **RAG chatbots**, intelligent multi-agent systems, and NFT-to-apparel generation pipelines that **automated mockup creation**, cutting design cycles from hours to minutes.

Skills

Programming Languages: Python, C/C++, Java, C#, Go, Bash, SQL

Machine Learning & GenAI: PyTorch, Transformers, Diffusers, LoRA/PEFT, xFormers, FAISS, CLIP, DINOv2, SAM

Large Language Models: Experience with state-of-the-art open-source and proprietary LLMs; skilled in prompting, RAG, fine-tuning, and evaluation; **LangChain, LangGraph, Hugging Face**

Computer Vision & Image Processing: Stable Diffusion (Playground v2.5/SDXL), Real-ESRGAN, SwinIR

Deployment & Data Engineering: FastAPI, WebSocket, Redis, Celery/RQ, ONNX, TensorRT

Cloud & MLOps: Azure (VMs, Container Apps, Marketplace), AWS (g5.4xlarge/g5.8xlarge), Docker, systemd, CI/CD

Web Scraping & Data Collection: Apify, Selenium, BeautifulSoup

Developer Tools: Git/GitHub Actions, Linux/Ubuntu

Soft Skills: Problem Solving, Self Learning, Teaching, Communication, Teamwork, Leadership

Professional Experience

AI/ML Engineer

VEEM Solutions (including Shrwd.ai project)

Jeddah, Makkah, Saudi

Arabia (Remote)

Sep 2024 – Present

- Designed and deployed a **large-scale multi-agent system** integrating a rich toolset, dynamic templates, and a **RAG knowledge layer** for intelligent orchestration, autonomous task execution, and context-grounded reasoning.
- Within the agent platform, delivered multilingual **conversational assistants** powered by FAISS + MiniLM retrieval, significantly improving customer support efficiency and response quality.
- Architected and implemented the **Shrwd.ai Brand Analysis & Tracking Platform**, the next evolution of brand intelligence—combining crawlers (Apify/Selenium/BeautifulSoup), SEO/engagement metric computation, and interactive dashboards for actionable insights.
- Engineered scalable **FastAPI/WebSocket pipelines** with Redis caching, Dockerized services, and **Azure** deployments (VMs & Container Apps) with CI/CD workflows and systemd automation.

AI Engineer (Part-time)

DigiTee.io

Cayman Islands (Remote)

Nov 2024 – Present

- Led the development of **NFT Wear AI**, fine-tuning Stable Diffusion with LoRA to automatically generate apparel mockups directly from NFT designs.
- Contributed to the **Mintora project**, building AI-driven design pipelines and interactive features such as generative design creation using Playground v2.5/SDXL, conversational chat integration, **background removal (rembg)**, and **high-resolution upscaling (Real-ESRGAN/SwinIR)**.
- Designed and deployed the **Audira multi-agent platform**, architecting intelligent agents with memory, STT (faster-whisper), TTS, and advanced orchestration to deliver multilingual voice/text onboarding and interaction.

- Managed **deployment to cloud servers**, configuring and maintaining **AWS g5.4xlarge/g5.8xlarge GPU instances**, handling scalability, worker orchestration, and load balancing for high-concurrency workloads.
- Implemented **GPU memory optimization** with xFormers and gradient checkpointing, reducing OOM errors and enabling stable concurrent inference across multiple workers.

Freelance AI/ML Engineer

Remote

Independent Projects

2024 – Present

- Delivered multiple **RAG-based solutions**, integrating FAISS/Hugging Face embeddings with FastAPI and WebSocket APIs for contextual Q&A systems.
- Designed and deployed custom **multi-agent architectures**, combining LangChain, LangGraph, and external tools for intelligent automation workflows.
- Built speech-driven applications with **STT (faster-whisper)** and **TTS (multi-lingual synthesis)** pipelines, enabling real-time voice assistants.
- Collaborated with clients on end-to-end AI projects, from model selection and fine-tuning to deployment on cloud infrastructure (**Azure, AWS, SaladCloud**).

AI Instructor

Asyut, Egypt

USAID (Ahmed Diefalla International Applied Technology School)

Oct 2024 – Apr 2025

- Taught courses in **Software Engineering, Python programming, Artificial Intelligence, and Robotics**.
- Guided students through hands-on projects and real-world applications, fostering innovation and practical problem-solving.
- Developed structured learning modules that improved students' technical and analytical skills.

Programming Content Creator (Part-time)

Karabük, Turkey (Remote)

Tweet Kod

Mar 2024 – Mar 2025

- Developed **educational programming content** for kids and teens using Python, including practical applications, projects, and chapter-end quizzes.
- Mentored students on graduation projects, improving their problem-solving and coding confidence.
- Created engaging lesson materials tailored for different learning levels.

Programming & AI Instructor (Part-time)

Cairo, Egypt (Remote)

Future Kids, Codemy, Code & Barmja

Jan 2024 – Dec 2024

- Delivered courses in **Scratch, Python, Java, HTML, AI, and Pygame** for children and young adults.
- Designed and tailored lesson plans to foster creativity, problem-solving, and programming confidence across diverse age groups.
- Led students through hands-on projects that combined technical skills with creativity and critical thinking.

Selected Projects

Eyes Mate — AI Application to Help the Blind (Graduation Project)

[GitHub Link](#) 

- Developed an assistive AI system enabling visually impaired users to interact with their environment using speech and computer vision.
- Implemented key features including object recognition, distance measurement, text reading, currency recognition, and face recognition with real-time voice feedback.

Brain Tumor Detection API

[GitHub Link](#) 

- Built a deep learning pipeline for brain tumor detection from MRI scans using CNN models.
- Deployed as a RESTful API, enabling easy integration into medical applications.

Generative AI Models

[GitHub Link](#) 

- Implemented and experimented with state-of-the-art generative AI architectures, including transformer and diffusion-based models.
- Focused on **text-to-image generation** and **image-to-text understanding** for creative and practical AI applications.

SmartAI Customer Support

[GitHub Link](#) 

- Designed an AI-powered customer support assistant leveraging retrieval-augmented generation (RAG).
- Improved customer response quality with contextual answers and multilingual support.

Deep CUDA — GPU Acceleration from Scratch

[GitHub Link](#) 

- Developed an advanced implementation of a **Convolutional Neural Network (CNN)** in CUDA C++ to classify images from the MNIST dataset.
- Built custom CUDA kernels for forward and backward propagation, matrix multiplication, convolution, and memory management, without relying on high-level frameworks.
- Optimized GPU performance through **thread/block management, shared memory usage, and parallel reduction techniques**, achieving efficient large-scale parallel computation.
- Demonstrated strong foundations in **low-level GPU programming**, bridging the gap between algorithm design and hardware acceleration for AI workloads.

Education

B.Sc. in Computer Science









Faculty of Computers and Information, Assiut University

Assiut, Egypt

2020 – 2024

- **Relevant Coursework:** Machine Learning, Deep Learning, Computer Vision, Parallel Computers

Certifications

- | | |
|---|----------|
| ◦ Generative AI with Large Language Models — Coursera
Certificate Link  | Apr 2025 |
| ◦ Advanced Learning Algorithms — DeepLearning.AI
Certificate Link  | Sep 2023 |
| ◦ Deep Learning with PyTorch: Image Segmentation — Coursera
Certificate Link  | Sep 2023 |
| ◦ Empedia Competitive Programming Competition — Empedia
Certificate Link  | Aug 2023 |
| ◦ Mathematics for Computer Science — BIO CODE, Assiut University
Certificate Link  | Aug 2023 |
| ◦ Supervised Machine Learning: Regression and Classification — DeepLearning.AI
Certificate Link  | Mar 2023 |
| ◦ Python Programming Basics — MaharaTech
Certificate Link  | Feb 2023 |
| ◦ Database Fundamentals — MaharaTech
Certificate Link  | Feb 2023 |

Publications & Research

- **AI Traffic Violation in Smart Cities and Flying Vehicles** — Research on AI-driven real-time monitoring, detection, and enforcement to enhance intelligent transportation systems and urban mobility.

Languages

Arabic (Native) English (Professional)