



Hazem Hassan

Senior Machine Learning Engineer (CV | GenAI | Edge AI)

Target Roles: Senior ML Engineer | ML Technical Lead

AI Engineer with **9+ years** shipping AI products in **CV/OCR, video analytics, edge AI, LLM Multi-RAG, and diffusion-based image generation/editing**. Strong in production deployment (cloud + Jetson), scalable ML APIs, and end-to-end ML ownership from modeling to monitoring.

Specialties: Edge AI (Jetson), Real-time CV pipelines, OCR, Multi-Object Tracking, Action Recognition, Multi-RAG systems, Diffusion image generation/editing, ML APIs, MLOps, Cloud deployment.

Email: hazem.mamdouh.fekry@gmail.com

Portfolio: hazemmamdouh.github.io

Phone: +201000912694

Location: Egypt (Open to Remote/Relocation)

LinkedIn: linkedin.com/in/hmamdouh

PROUD ACHIEVEMENTS

- Delivered real-time video analytics on **NVIDIA Jetson**, enabling low-latency production inference for tracking, counting, and event detection.
- Led end-to-end delivery of **eKYC identity verification systems** (OCR, face match, liveness), shipping multiple production releases.
- Built **Multi-RAG** pipelines for LLM applications (multi-source retrieval, reranking, evaluation), improving response quality and reducing hallucinations.
- Designed **diffusion-based** image generation and editing workflows for controllable synthesis and transformation.
- Won **1st Place** in **MGB-5 (2019)** and **RASM (2018)** competitions.

CORE SKILLS

ML/DL	Model Training, Model Fine-Tuning, Feature Engineering, Hyperparameter Tuning, designing a hybrid objective function, Evaluation Metrics, Error Analysis	GenAI & LLMs	Large Language Models (LLMs), Multi-RAG, Retrieval-Augmented Generation, Embeddings, Reranking, Chunking Strategies, Prompt Engineering, LLM Evaluation	Programming	Python, SQL
Computer Vision & NLP	Object Detection, Segmentation, OCR, Video Analytics, Multi-Object Tracking, Action Recognition, Anomaly Detection, Face Recognition, Liveness Detection, Image Processing	Diffusion Models	Image Generation, Image Editing, Controllable Image Synthesis	Frameworks	PyTorch, TensorFlow, Scikit-learn
Deployment & MLOps	Model Serving, FastAPI, REST APIs, Docker, Kubernetes, CI/CD, Model Optimization, ONNX, TensorRT, Inference Performance Tuning	Cloud & Tools	GCP, AWS, Git, Linux	Edge AI	NVIDIA Jetson, Real-time Inference, Low-latency Video Pipelines
Leadership	Technical Leadership, System Design, Cross-functional Collaboration, Mentoring, Code Reviews				

WORK EXPERIENCE

Senior ML Engineer

Beyond Limits

03/2024 – Present

A pioneering AI company with a unique heritage for the US space program

Achievements/Tasks

EMEA Team

- Deployed real-time video analytics on **NVIDIA Jetson edge devices** across **10+ camera feeds / 5+ sites**, sustaining **25–35 FPS** with **<150 ms** latency for tracking, counting, and behavior analytics.
- Delivered embedded analytics for **Single/Double Flow Easier Gates**, achieving flow accuracy by **~94%** and attained tracking ID switches by **~14%** through optimized detection + MOT tuning.
- Built **PPE compliance detection** (Aramco), achieving **~94% precision / ~90% recall**, reducing false alarms by **~35%** via post-processing and threshold calibration.
- Developed CCTV **child violence detection** (spatio-temporal modeling), achieving **F1 ~84%**, using only **10-mins** of training data, and cutting annotation time by **~80%** using a semi-automated labeling workflow reaching **F1 ~80%** on the same training data.
- Implemented dense-scene people counting/crowd distribution, reducing counting error (MAPE) by **~43%** and supporting **7+ concurrent streams**.
- Built **Multi-RAG pipelines** (multi-source retrieval + reranking), improving **Recall@10** by **~20%** and reducing hallucinations by **~30%** using evaluation-driven iteration.

Tech Stack: PyTorch, OpenCV, NVIDIA Jetson, Python, FastAPI, Docker, Kubernetes, ONNX/TensorRT, Vector Search

WORK EXPERIENCE

Technical Lead

IDefy (Digital Identity/eKYC)

10/2022 - 02/2024

A Spin-Off from RDI, Specialized for Digital Identity Solutions for Onboarding Services and eKYC enabler

Achievements/Tasks

- Led digital identity onboarding products (SaaS/SDK/web/mobile), delivering **3 production deployments** end-to-end.
- Managed a cross-functional team (**8-12 engineers**) and built APIs for OCR, face matching, liveness, and identity verification, processing **10K+ sessions/month** at **99.9% uptime**.
- Developed an automated system to be adapted to any structured documents, getting few samples and reaching **~95% of recognition accuracy**.
- Released **IDefy eKYC App** and **GAMA Visitor Management**, improving verification success rate by **~20%** via quality tuning and validation workflows.

Giza, Egypt

Senior ML Engineer (Computer Vision)

RDI

01/2021 - 10/2022

Leading company related to Arabic Natural Language Processing using Machine Learning and AI solutions

Achievements/Tasks

Giza, Egypt

- Led the delivery of the **Sotoor Arabic OCR System**, achieving **~12 WER**.
- Built document OCR pipeline modules (layout analysis, segmentation, font detection, recognition, denoising, correction), reducing OCR error by **~7% on noisy scans**.
- Adapted the K2 Toolkit for OCR training/decoding, improving training efficiency by **~25%** and gaining **~4% WER**, boosting sequence recognition stability.

SELECTED PROJECTS

Agentic Agent Call Center

12/2025 - Present

- Built an **agentic AI call-center assistant** using LLM tool-calling and Multi-RAG retrieval to automate intent handling, answer generation, and CRM/ticketing actions.
- Designed a **reliable conversation orchestration layer** with guardrails, escalation-to-human routing, and quality evaluation to ensure safe and consistent customer support automation

Fashimi App

05/2025 - Present

- Built an AI try-on platform using **LLMs for recommendation + prompt generation** and **diffusion models for avatar generation and photorealistic try-on rendering**.
- Developed a recommendation engine leveraging mood/occasion/climate/user preferences to produce personalized outfit prompts.

Valorant Overlay

01/2022 - 06/2022

- Designed a real-time multi-threaded system extracting gameplay data from Valorant for esports competition overlays.
- Implemented object detection, pattern recognition, and color analysis, achieving a response latency of **50-72 msec**.

PUBLICATIONS

Arabic Multi-Genre Broadcast Speech Recognition

12/2019 - Examination Paper

IEEE ASRU 2019

Prognostic AI for Post-Operative Infections

08/2023 - 11/2023 - Conference Paper

FCI Annual Conference

QuranScript Verification System

11/2016 - Conference Paper

Sixteenth ESOLE Conference of Language Engineering

EDUCATION

Bachelor of Science

Electrical Electronics and Communication Engineering

Cairo University, 2016, with **GPA: 3**

VOLUNTEER EXPERIENCE

Project Mentor & Judge – Information Technology Institute (ITI)

11/2021 - 01/2022

Mentored graduation project in Math Formula Detection and Recognition, advising on CV algorithms and system design.

HOBBIES

• Swimming

• Gaming

• Padel