Hazem Hassan

Cairo, Egypt hazem.mamdouh.fekry@gmail.com (+20) 100-0912-694 in/hmamdouh hazemmamdouh.github.io

PROFESSIONAL SUMMARY

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

ROLES AND SPECIALITIES

Roles: Senior ML Engineer, AI Technical Lead.
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.

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.

WOEK EXPERIENCE

Senior ML Engineer

Beyond Limits — March 2024—Present, Remote, EMEA

- 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

Technical Lead

IDefy (Digital Identity/eKYC) — October 2022—February 2024, Giza, Egypt

- 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 a few samples and reaching ~95% of recognition accuracy.
- Released the IDefy eKYC App and GAMA Visitor Management, improving verification success rate by ~20% via quality tuning and validation workflows.

Senior ML Engineer (Computer Vision)

RDI — January 2021—October 2022, 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.

PROJECT

Agentic Agent Call Center

Co-Founder

- 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
Lead Engineer
- 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
Developer
- 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.

EDUCATION

Bachelor of Science in Engineering
Cairo University – Electrical Electronics and Communication Engineering · 2016

SKILLS

Programming: Python, SQL.
Frameworks: PyTorch, TensorFlow, Scikit-learn.
ML/DL: Model Training, Model Fine-Tuning, Feature Engineering, Hyperparameter Tuning, designing a hybrid objective function, Evaluation Metrics, Error Analysis.
Computer Vision & NLP: Object Detection, Segmentation, OCR, Video Analytics, Multi-Object Tracking, Action Recognition, Anomaly Detection, Face Recognition, Liveness Detection, Image Processing.
GenAI & LLMs: Large Language Models (LLMs), Multi-RAG, Retrieval-Augmented Generation, Embeddings, Reranking, Chunking Strategies, Prompt Engineering, LLM Evaluation.
Diffusion Models: Image Generation, Image Editing, Controllable Image Synthesis.
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.

PUBLICATIONS

Arabic Multi-Genre Broadcast Speech Recognition
Examination Paper, December 2019 – IEEE ASRU 2019

Prognostic AI for Post-Operative Infections
Conference Paper, November 2023 – FCI Annual Conference

QuranScript Verification System
Conference Paper, November 2016 – Sixteenth ESOLE Conference of Language Engineering

HOBBIES

Swimming Gaming Padel