

Hossam Elsherbiny

Computer Vision Engineer

Cairo, Egypt | hossam.elsherbiny95@gmail.com | +20 114 721 1427 | hossam-elsherbiny.me
github.com/H-Elsherbiny | linkedin.com/in/h-elsherbiny

Professional Summary

Computer Vision Engineer with hands-on experience developing deep learning pipelines for object detection, multi-object tracking, image classification, geometric transformation, and generative vision. Skilled in Python, PyTorch, OpenCV, YOLO, and Vision Transformers, with experience fine-tuning and evaluating models for real-world image and video applications.

Experience

AI Engineer, Freelance Oct 2024 – Present

- Fine-tuned the FLUX.1-dev text-to-image model using LoRA on NVIDIA A100 GPU, achieving high-fidelity custom character generation.
- Built AI pipelines for Egyptian Arabic speech recognition, NER extraction, and video-based gesture detection using Whisper, MARBERT, GPT-4o-mini, MediaPipe, and OpenCV.

AI Engineer, Acrow – Cairo Nov 2024 – Oct 2025

- Designed time-series forecasting pipelines for 3,000+ SKUs using SARIMA, Prophet, XGBoost, CatBoost, LightGBM, and neural networks, reducing forecast error from 80% to 55%.
- Engineered a GUI-based forecasting application that automated the end-to-end pipeline, reducing execution time from 3 days to 1.5 hours and increasing profit by 10.5%.

Education

9-Month Professional Program – AI & ML Track, Information Technology Institute (ITI) Oct 2025 – Present

B.Sc. in Computer Engineering – Faculty of Engineering, Zagazig University Oct 2020 – Jul 2025

Projects

Football Analyzer [GitHub](#) | [Portfolio](#)

- Developed an end-to-end football analytics pipeline using fine-tuned YOLOv26 and ByteTrack for player, referee, and ball detection and multi-object tracking, achieving 94.6% mAP@0.5.
- Implemented homography-based pitch calibration and CIELAB K-Means team classification, achieving <5% projection error and 98.2% classification accuracy for tactical and performance analytics.

CareFlow Clinical Assistant [GitHub](#) | [Portfolio](#)

- Deployed a 10-microservice multimodal clinical AI platform integrating MedGemma 1.5 VLM, MedSAM 2.0, and Arabic ASR to automate radiology analysis and reduce physician documentation time by up to 40%.
- Designed a computer vision pipeline for medical image interpretation, ROI segmentation, and layout-aware document OCR, achieving 96% precision and 98% recall for automated PII redaction and structured clinical information extraction.

DeepFM Architecture Implementation [GitHub](#) | [Portfolio](#)

- Reproduced the *DeepFM Paper* from scratch for Click-Through Rate (CTR) prediction on the Criteo dataset (~45M samples, 39 features), reproducing the published benchmark results (AUC $\approx$ 0.801, LogLoss $\approx$ 0.456).
- Conducted 6 ablation studies and optimized the model for high-dimensional, imbalanced data, improving performance to AUC $\approx$ 0.802 and F1-score $\approx$ 0.56.

Technical Skills

Programming Languages: Python, SQL, C++

ML Frameworks: PyTorch, TensorFlow, Scikit-learn

LLMs & Agentic AI: Transformers, Hugging Face, LangChain, LangGraph, LangSmith, RAG

Computer Vision (CV): OpenCV, YOLO, Vision Transformers (ViT), VAEs, GANs, Diffusion Models

MLOps & Cloud: Docker, Kubernetes, MLflow, Git, AWS

Big Data & Databases: PySpark, Kafka, PostgreSQL, MongoDB, ChromaDB

Backend: FastAPI, REST APIs

Additional Information

Military Status: **Exempted** (Certificate available upon request).