

YOUSSEF KHALIFA

Nasr City, Cairo, Egypt • +20 102 244 5304 • youssekhalifa458@gmail.com • [LinkedIn](#)

SUMMARY

AI Engineer & Data Scientist with **3 + years** of experience designing, deploying, and scaling AI solutions for computer-vision and NLP workloads. Delivered production-grade facial-recognition, license-plate detection, and BERT-based complaint-classification systems that reduced manual effort by **up to 90 %** and cut response times by **75 %**. Proficient in Python, TensorFlow, FastAPI, Docker/ Kubernetes, MCP server integration, and LLM-orchestrated agent frameworks.

EDUCATION

Ain Shams University – B.Sc. in Computer Science & Information Systems, 2024
Major: Computer Science GPA: 3.05

WORK EXPERIENCE

AI Engineer | 51Talk

OCT 2025 – PRESENT

- Identified **12** high-value AI use cases and integrated them into core business workflows, increasing process automation coverage by **30 %**.
- Built and deployed a **personalized recommendation engine** with TensorFlow and Azure ML, lifting user engagement by **15 %**.
- Implemented **MCP server-based model serving** in Docker containers, reducing model latency from **350 ms to <200 ms** (≈ 43 % improvement).
- Orchestrated LLM-driven chat assistants via LangChain, handling **5,000+** concurrent user queries with **98 %** satisfaction.
- Added **AWS SageMaker** support for model training and **GitHub Actions** CI/CD pipelines, cutting deployment time by **40 %**.

Data Scientist | Emaar Misr

OCT 2024 – SEP 2025

AI-Powered Business Intelligence Chatbot

- Co-designed a chatbot that translated natural-language queries into SQL, generating **200+** interactive visualizations per month.

- Developed an end-to-end data pipeline (SQLAlchemy → Apache Spark) that processed **10 TB** of raw data weekly, cutting ETL time by **45 %**.

ENTERPRISE SURVEILLANCE INTELLIGENCE PLATFORM

- Served as sole architect; designed micro-service architecture with FastAPI, Docker, and Kubernetes, achieving **99.5 %** uptime.
- Authored **75 %** of the codebase, including a face-recognition engine (InsightFace) and license-plate detector, reaching **85 %** accuracy and **<200 ms** latency.
- Optimized FAISS similarity search from $O(n)$ to $O(\log n)$, enabling **30 FPS** matching against a **50k+** resident database.
- Standardized detection payloads using JSON schema + base64 encoding, reducing client-side integration effort by **60 %**.
- **Impact:** Cut security response time by **75 %** and eliminated **90 %** of manual monitoring tasks.

NLP Complaint Classification System

- Built a BERT-based classifier that processed **258 k** tickets across **158** categories, attaining **78 %** accuracy.
- Integrated BiLSTM with BERT embeddings (1.5 M parameters) and used Optuna (50+ trials, 8-worker parallelism) for hyper-parameter tuning, reducing manual routing by **90 %**.
- Applied SMOTE for class imbalance and linked the model to Snowflake for multilingual preprocessing.
- **Impact:** Automated **60 %** of ticket routing, decreasing average handling time from **12 min to 4 min**.

Power BI Enterprise Dashboards

- Delivered 5 cross-functional dashboards (Complaints & CSAT, Delayed Tasks, SLA Adherence, P&L, VIP Customer Insights) used by **200+** stakeholders for real-time decision making.

CERTIFICATIONS & COURSES

- **Sample-Based Learning Methods** – University of Alberta & Coursera, Sep 2025
 - **Machine Learning** – DeepLearning.ai, Stanford & Coursera, Mar 2025
 - **TensorFlow** – DeepLearning.ai & Coursera, Feb 2025
 - **IBM Data Science Professional Certificate** – Coursera, Feb 2025
 - **Microsoft Azure Machine Learning** – Coursera, Jan 2025
-

PROJECTS

X-Ray Bone Abnormality Detection (MURA) – 2024

- Developed a two-stage CNN pipeline; single-view model achieved **82 %** accuracy, multi-view model reached **85 %** on the MURA dataset.

Neural Network & Transformer from Scratch – 2023

- Implemented a full-stack neural network and transformer architecture without external libraries, demonstrating deep understanding of back-propagation and attention mechanisms.

SKILLS

- **Programming:** Python (used for recommendation engine, BERT classifier), Java, C++, C#, JavaScript, Scala
 - **AI/ML:** Deep Learning, Computer Vision, NLP, Transformers, TensorFlow, PyTorch, Scikit-learn, Hugging Face (applied in face-recognition and complaint classification)
 - **MLOps & Infrastructure:** Docker, Kubernetes, Azure ML, AWS SageMaker, MCP server deployment, FastAPI, Micro-service architecture, LangChain, PromptFlow, Git/GitHub, GitHub Actions CI/CD, MLflow (model tracking)
 - **Data Engineering:** SQL, SQLAlchemy, Apache Spark, Snowflake, Data pipeline optimization, ETL design (built 10 TB/week pipeline)
 - **Analytics & Visualization:** Power BI (5 enterprise dashboards), Excel, DAX, Tableau (basic)
 - **Soft Skills:** Presentation, Communication, Leadership, Teamwork, Adaptability, Time Management
-