

ISLAM KASSEM

AI / ML Engineer · LLM & GenAI Specialist · MLOps · Computer Vision

Riyadh, Saudi Arabia | iqasem4444@gmail.com | [\(+966\) 538 562 171](tel:+966538562171) | [linkedin.com/in/elcaiseri](https://www.linkedin.com/in/elcaiseri)

PROFESSIONAL SUMMARY

AI/ML Engineer designing, fine-tuning, and deploying production-grade AI systems, including LLM-powered agents, RAG pipelines, Whisper ASR, and multimodal computer vision solutions. Kaggle Master (Top 1% globally) and Upwork Top-Rated Freelancer with a 100% Job Success Score across 50+ international projects. Strong expertise in LoRA/QLoRA fine-tuning, inference optimization, and scalable MLOps systems. Currently architecting enterprise AI solutions at Master Works / Nabeh, contributing to Saudi Vision 2030 digital transformation initiatives.

CORE SKILLS & TECH STACK

AI / ML: LLMs, RAG, Whisper ASR, LoRA / QLoRA Fine-Tuning, Computer Vision, Prompt Engineering, Semantic Search, AI Agents, NLP, Deep Learning, Multimodal AI, Recommendation Systems

Models: GPT, LLaMA 3.x, Mistral, Gemini, Claude, OpenAI models, Whisper (large-v3), CLIP, BLIP, Sentence-Transformers, vLLM, Ollama

Frameworks: PyTorch, TensorFlow, Hugging Face Transformers, LangChain, scikit-learn, OpenCV, XGBoost, SpaCy, NLTK, FastText, FastAPI

Languages: Python (advanced), SQL, R, Bash, NoSQL — MongoDB, Cassandra

Cloud / MLOps: AWS, Azure, GCP (Vertex AI), Docker, Kubernetes, CI/CD, GitHub Actions, MLflow, Weights & Biases, A/B Testing, Model Monitoring

Vector / Search: FAISS, Pinecone, Weaviate, Elasticsearch, Apache Kafka, REST APIs, gRPC

PROFESSIONAL EXPERIENCE

AI Engineer · Master Works — Nabeh AI, Riyadh, SA

Nov 2025 – Present

- Architected and deployed enterprise-grade AI systems including RAG pipelines (FAISS/Chroma), LLM-powered chatbots, and autonomous agents for large-scale Saudi enterprise use cases.
- Led on-premise LLM deployments (LLaMA, Qwen) on GPU-based Kubernetes infrastructure, enabling secure, data-sovereign GenAI solutions for privacy-critical environments.
- Built production-ready Speech-to-Text (Whisper ASR) pipelines processing large-scale audio data, integrated with NLP extraction and downstream analytics workflows.
- Designed end-to-end MLOps pipelines with MLflow, Weights & Biases, and GitHub Actions CI/CD, improving experiment tracking, deployment reliability, and model delivery efficiency.

AI / ML Engineer · Velents, Riyadh, SA

Jul 2024 – Aug 2025

- Productionized Generative AI systems (LLMs, RAG, Whisper ASR) to automate recruitment and workforce management workflows for government and enterprise clients.
- Built an AI meeting assistant using Whisper + LLMs for automated transcription, summarization, and action-item extraction, reducing manual documentation effort.
- Developed semantic resume-to-job matching engine using embeddings and RAG, improving candidate–job alignment across recruitment and education domains.
- Designed and deployed video interview intelligence and AI-driven fraud detection systems, integrating facial recognition, sentiment analysis, and document processing for enterprise HR workflows.

Machine Learning Engineer · TwentyToo, Cairo, EG

Aug 2023 – Jul 2024

- Designed and deployed an AI-powered fashion search engine using transformer embeddings and FAISS-backed ANN retrieval, significantly improving product discoverability.
- Implemented auto-tagging and image retrieval systems using CNNs and transfer learning, streamlining fashion catalog management at scale.
- Built recommendation pipelines combining collaborative filtering and embedding-based retrieval, driving measurable gains in user engagement metrics.

AI & ML Engineer · Upwork, Remote

Apr 2021 – Aug 2023

- Delivered 50+ end-to-end AI/ML projects spanning NLP, computer vision, and predictive analytics for global clients; achieved and maintained 100% Job Success Score and Top-Rated status.
- Architected and deployed scalable ML APIs, ETL pipelines, and inference services on AWS, GCP, and Azure for diverse business verticals.
- Fine-tuned pre-trained NLP models for domain-specific classification, NER, and semantic search tasks; built reproducible experimentation frameworks using MLflow.

Data Scientist & Community Contributor · Kaggle, Remote

Oct 2018 – Apr 2021

- Competed in 15+ Kaggle competitions, ranked in global Top 1% as Kaggle Code Master.
- Published 15+ high-impact educational notebooks on NLP, CV, and ML fundamentals.

FEATURED PROJECTS

OpenFPL Scout AI | Python · ML · Time Series · Predictive Analytics · Live Data APIs

- AI-powered Fantasy Premier League scout using ML time-series models and real-time data to predict player performance and automate team-selection decisions. [kassem.dev/open-fpl]

AI Meeting Assistant Agent | Whisper ASR · LangChain · LLMs · Python · Automation

- End-to-end autonomous agent that transcribes meeting audio via Whisper, generates structured summaries and action items with LLMs, and integrates with workflow automation.

Resume–Role Semantic Matching (RAG) | RAG · FAISS · Embeddings · Sentence-Transformers

- Recruitment-grade semantic search system that matches candidate resumes to job roles using dense vector retrieval and re-ranking, deployed for government hiring workflows.

LangChain MCP Adapter — Open Source | LangChain · MCP · Python · LLM Agents

- Contributed to the LangChain MCP adapter ecosystem, extending open-source tooling for LLM agent orchestration and cross-service integration. [github.com/elcaiseri]

EDUCATION

M.Sc. in Data Science

Jan 2024 – Present

Cairo University — Faculty of Graduate Studies for Statistical Research

B.Sc. in Electrical & Communication Engineering

Sept 2016 – Jul 2021

Port Said University — Faculty of Engineering

CERTIFICATIONS

- Deep Learning Specialization — DeepLearning.AI
- Analyze Datasets & Train ML Models using AutoML — Amazon Web Services

LANGUAGES

- **Arabic** (Native)
- **English** (Fluent)
- **Spanish** (Elementary)