Buthaina Esam Mohamed

Giza, Egypt|+201020636444| bosinaesam78@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

Summary

Computer Science graduate passionate about building practical ML and NLP systems. Experienced with transformer models, hybrid retrieval, and local deployment using FastAPI, Docker, and MLflow. Skilled in quantization, model tracking (DagsHub), and workflow automation with n8n. Effective communicator with teaching and leadership experience — seeking an AI/ML role to apply and expand real-world engineering skills.

Education

BSc in Computers and Artificial Intelligence   Helwan University, Cairo

GPA: Excellent with Honors   Sept 2021 – June 2025

Technical Skills

- **ML & Deep Learning:** PyTorch, TensorFlow, Scikit-learn, Keras
- **NLP & CV**: Hugging Face Transformers, spaCy, NLTK, OpenCV, LangChain, LiveKit
- **Deployment & Optimization**: FastAPI, Streamlit, Docker, MLflow, DagsHub,Quantization, Azure, Git ,n8n
- **Data & Visualization**: Pandas, NumPy, Seaborn, Plotly
- **Programming Languages**: Python (main), Java, C, SQL, HTML/CSS, JavaScript

Work Experience

AI Superheros   Remote

AI & Automation Intern   Sept 2025 - Oct 2025
- Built two automation projects using n8n: an AI WhatsApp chatbot and a RAG-based chatbot integrating workflow logic and external APIs.

Elevvo Pathways   Remote

NLP Intern   Sept 2025 - Oct 2025
- Completed 4 NLP projects covering text analysis, summarization, QA systems, and topic modeling.
- Improved summarization accuracy by +5.2 ROUGE-1 and compared BERT vs DistilBERT on SQuAD.

ischool   Remote
DEMI Coding Instructor   March 2025- April 2025
- Taught 6th grade students the fundamentals of coding and game development using Godot and GDScript.

Project Experience

Botify – AI Chatbot Builder (Graduation Project) *[GitHub](#)* | *[Demo](#)* | *[Documentation](#)*
Transformer Models | Hybrid Retrieval | Quantization

- Designed and implemented the AI pipeline using 4 transformer models: BART (intent detection), BERT (NER), RoBERTa (sentiment), and LLaMA (response generation), selected after testing multiple models for accuracy, speed, and deployability.

- Built a document preprocessing pipeline for PDFs, DOCX and TXT with chunking based on semantic boundaries, lemmatization, and tokenization using NLTK and spaCy.
- Developed a hybrid retrieval system combining BM25, FAISS (MiniLM embeddings), and KeyBERT topic matching to retrieve the most relevant context chunks.
- Optimized runtime by quantizing local models (intent, NER, sentiment) to INT8; while using LLM response generation via API, with real-time fallback logic based on sentiment, repetition, and confidence.

LiveKit Voice Agent with RAG Integration *GitHub*
LiveKit | Gemini Live | RAG | FAISS | Sentence-Transformers

- Built a real-time voice AI agent using LiveKit + Gemini Live with end-to-end audio streaming.
- Added RAG via FAISS + MiniLM embeddings, triggered through Gemini function calling for grounded answers.
- Delivered low-latency, knowledge-aware voice responses from a custom company knowledge base.
- Created a React frontend using LiveKit Web SDK for room joining and voice interaction.

Loan Approval Prediction API *GitHub*
FastAPI | XGBoost | Scikit-learn | Docker | MLflow
- Built a production-ready FastAPI service for real-time loan approval prediction using an optimized XGBoost model (95.6% accuracy).
- Performed full data preprocessing: correlation analysis, feature removal, label encoding, outlier capping, and MinMax scaling on a 4,269-record dataset.
- Deployed the model as a REST API and Swagger docs; containerized the service with Docker and tracked experiments using MLflow.

Arabic-to-English Neural Machine Translation *GitHub* | *presentation*
MarianMT | Hugging Face Transformers | BLEU/METEOR | Streamlit App

- Fine-tuned the Helsinki-NLP/opus-mt-ar-en model on the Global Voices dataset (56k cleaned sentence pairs) to improve Arabic-English translation quality.
- Applied thorough data preprocessing, tokenization, and custom PyTorch Dataset construction for efficient training.
- Trained with Seq2SeqTrainer (3 epochs, batch size 32), achieving BLEU score 40.4 and METEOR 0.70 on test data and deployed a user-friendly Streamlit app locally.

Courses and Certifications

- Generative AI Diploma By DEPI   June 2025 – Present
- ATEFL (Cairo University) – Score: 567 / 700 (Equivalent to B2–C1)   November 2025
- Machine learning for Data Analysis By NTI   March 2025 – June 2025
- Microsoft ML Diploma By DEPI *Link*   April 2024 – October 2024

Volunteering Experience

Ministry of Youth and Sports, Egypt, YLY

Coordination Member → Planning Leader → Trainer   August 2022 – June 2025
- Trained 200+ youth in soft skills and planning; coordinated national events with the Ministry.