

AHMED GONGI

AI Engineer — Generative AI — LLM & RAG Specialist

Monastir, Tunisia (Open to International Relocation)

✉ contact.gongiahmed@gmail.com —  LinkedIn —  GitHub —  Portfolio —
+216 94853120

PROFESSIONAL SUMMARY

AI Engineer with hands-on experience in Generative AI, Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG) systems. Proven ability to design, fine-tune, and deploy production-grade AI solutions integrated with enterprise environments. Experienced in backend development (.NET, REST APIs), database integration (SQL, AS/400), and workflow automation. Strong background in machine learning, NLP, and computer vision with a focus on scalable and secure AI deployment.

PROFESSIONAL EXPERIENCE

AI Engineer

Tunisian Travel Service (TTS)

2025 – Present

- Designed and deployed a Generative AI virtual assistant using LLMs and Retrieval-Augmented Generation (RAG).
- Integrated WhatsApp Business API with AI backend services via n8n workflow automation.
- Developed and maintained backend services using .NET (C#) and REST APIs.
- Integrated AI pipelines with AS/400 and SQL databases for secure booking data retrieval.
- Implemented data validation and access control logic for sensitive customer information.
- Optimized API usage through rate limiting and API key rotation strategies.
- Administered COSWIN ERP system for stock and maintenance management.
- Generated operational and analytical reports using Crystal Reports.

Data Scientist

Smart Conseil

2025

- Developed predictive analytics solutions based on time series data.
- Leveraged NLP techniques to extract structured information from unstructured documents.
- Implemented computer vision models to automate visual tasks.
- Integrated and fine-tuned LLMs for business process automation.

LLM Engineer (Final Year Project)

Mobelite

2024

- Built and fine-tuned a LLaMA 3 model using LoRA and QLoRA techniques.
- Automated DevOps workflows using a customized LLM-based solution.
- Performed large-scale data preprocessing and dataset optimization.

Computer Vision Engineer Intern

SARTEX

2023

- Developed a real-time face detection system using FaceNet and OpenCV.
- Achieved 95% accuracy in employee recognition deployment.

ML Engineer Intern

SARTEX

2022

- Developed a defect prediction model reducing manufacturing quality issues.
- Performed feature engineering and exploratory data analysis.

TECHNICAL SKILLS

- **Generative AI & LLMs:** LLaMA, Mistral, Hugging Face Transformers, LoRA, QLoRA, PEFT, Prompt Engineering, RAG, LangChain
- **Machine Learning:** Supervised & Unsupervised Learning, XGBoost, LightGBM, CatBoost, Feature Engineering, Model Evaluation
- **Natural Language Processing:** Text Classification, Sentiment Analysis, Named Entity Recognition
- **Computer Vision:** CNNs, Face Recognition, YOLO, OpenCV, Transfer Learning
- **Backend Development:** .NET (C#), REST APIs, FastAPI, Flask
- **Databases:** SQL Server, MongoDB, AS/400
- **Deployment & DevOps:** Docker, n8n, Streamlit
- **Programming:** Python, SQL

SELECTED PROJECTS

CV Extractor with Open-Source LLMs

- Developed a CV information extraction system using LLaMA 3 and Mistral via Ollama.
- Processed scanned and text-based PDFs using OCR and evaluated performance with precision, recall, and F1-score.

Sentiment Analysis with Fine-Tuned LLaMA 3

- Fine-tuned an 8B LLaMA 3 model using QLoRA for financial sentiment classification (86.1% accuracy).

Industrial Production Forecasting

- Forecasted industrial time series using ARIMA, Prophet, and Chronos (best MAPE: 0.009).

Facebook Scraper & Toxicity Analyzer

- Built a FastAPI service integrating Detoxify and MongoDB for automated toxicity analysis.

EDUCATION

Engineering Degree in Artificial Intelligence and Data Science
EPI, Sousse

2021 – 2024

Preparatory Cycle for Engineering Studies
IPEIM, Monastir

2017 – 2021