MHEDHBI Hamza

+216 93 377 688 | mhedhbihamza11@gmail.com | linkedIn | portfolio | Ariana, Tunisia

AI Engineer building AI that solves real-world problems with LLMs, NLP, OCR, and modern AI infrastructure.
Building user-centric AI products that people actually use.

Experience

AI Engineer 10 / 2025 –
Poulina Group Holding

Project 1: Complaint Analytics Platform

- Built a multi-hotel complaint analytics platform.
- Implemented **role-based access control** with JWT authentication, for permission-based access.
- Implemented LLM-based **complaint classification** using LLaMA 3.1 8B with batching,and retry logic.
- Automated AI action plan generation (LLaMA 3.3 70B), reducing reporting time from **hours to seconds.**
- Built a **Kanban-style task management board** linking complaint to track tasks across lifecycle states.
- Containerized and deployed the platform on an Oxahost VPS using Docker Compose and Nginx.

Keywords: FastAPI, React, Groq, LLaMA, Data Pipeline, RBAC, JWT Authentication, Kanban, Docker, Nginx, VPS Deployment

Project 2: CleverTap Integration

- Unified **8 disparate systems** (2 PMS, events, VR venue, pool, timeshare, park, WiFi) across 5 hotels and 2 parks into a single CleverTap(**CRM**) view.
- Replaced manual CSV exports with **hourly real-time sync** via a fully automated ingestion pipeline.
- Designed an MD5 hash-ledger deduplication engine decoupled from unreliable DB timestamps to eliminate duplicates.
- Achieved **98–100% record coverage** via ETL logic handling dates, null/NaN sanitization, and E.164 phone normalization.
- Deployed **24/7 with zero downtime** on Ubuntu VPS.

Keywords: CleverTap, ETL, Deduplication, MD5, Data Pipeline, Multi-source Integration, E.164, Docker, Flask, Ubuntu VPS

AI Engineer Intern – OCR Pipeline Upgrade 02 / 2025 – 06 / 2025
BinitNS - Accelex

- Optimized SmolDocling's OCR performance using patching, overlapping, and MSER.
- Deployed Qwen-VL 2.5 on AWS SageMaker via Ollama, enabling real-time parsing.
- Engineered a custom Bash script to maintain Ollama's operational sustainability in production, ensuring **100%** uptime.
- Proposed a modern OCR pipeline upgrade for Accelex, **replacing legacy Tesseract-based systems with VLM-powered solutions.**

Keywords: OCR, Vision Language Models, SmolDocling, Qwen 2.5 VL, AWS SageMaker, Ollama, Bash, MSER, Bounding Box Detection

AI Engineer Intern – LegalBot 10 / 2024 – 02 / 2025
Approach AI

- Designed a pipeline for automated legal document preprocessing and OCR using Azure Blob Storage.
- Reduced Q&A pairs generation time by **90%** using GPT-4, Llama 3:8B, and Unsloth AI, with integration into Azure ML Studio.
- Built RAG with FAISS for efficient data retrieval.

Keywords: Azure Blob Storage, Azure ML Studio, OpenAI GPT-4, Semantic Search, Legal Document Parsing

AI Engineer Intern – French-Language HR Chatbot 06 / 2024 – 08 / 2024
SACEM Industries

- Developed a French-language chatbot using LLaMA3 8B fine-tuned with LoRA, enhanced with SpaCy for NER, for HR use.
- Implemented RAG (Retrieval-Augmented Generation) with FAISS for efficient data retrieval.
- Synchronized SharePoint lists with Excel Online using Power Automate for **real-time** data access.
- Built a real-time interface using Streamlit.

Keywords: French-language chatbot, LLaMA3 8B, LoRA fine-tuning, NER, RAG, FAISS, SharePoint, Excel, Power Automate, Streamlit.

AI Engineer Intern – Speaker Identification and Transcription System 06 / 2024 – 08 / 2024
Omnilink

- Applied noise reduction, normalization, and segmentation to prepare audio for optimal model performance.
- Combined pyannote/speaker-diarization with Whisper AI for accurate speaker segmentation and transcription.
- Utilized LLaMA2 for language understanding to classify speakers as employees or customers.
- Built a Flask-based interface for portability and scalability.
- The startup plans to bring the project to market.

Keywords: Audio preprocessing, noise reduction, normalization, segmentation, pyannote, Whisper AI, transcription, LLaMA2, speaker classification, Flask, startup, market-ready.

ML Intern – SMS Spam Detection System 07 / 2023 – 08 / 2023
Tunisie Telecom

- Designed and implemented an SMS spam detection system using advanced NLP and machine learning techniques.
- Trained and evaluated models including **Multinomial Naive Bayes, XGBoost, and fine-tuned BERT** to detect spam.
- Applied text preprocessing methods such as stopwords removal, stemming, and tokenization to clean and structure input data.
- Used feature extraction methods including **CountVectorizer, TF-IDF, and GloVe embeddings** to represent text data effectively.

Keywords: NLP, Machine Learning, Multinomial Naive Bayes, XGBoost, BERT, Text Preprocessing, CountVectorizer, TF-IDF, GloVe, Spam Detection, Feature Engineering.

Education

Engineering degree in Statistics and Data Analysis, Higher School of Statistics and Information Analysis 2022 — 2025
Pre-engineering degree (Maths and Physics), Preparatory Institute of Nabeul 2020 — 2022

Technical Skills

Programming: Python, JavaScript, R, Java, HTML, CSS

AI/ML Frameworks: PyTorch, TensorFlow, Scikit-learn, Hugging Face

NLP & LLMs: Transformers, SpaCy, NLTK, LLaMA, GPT-4, BERT, LoRA, Prompt Engineering

Web & Backend: FastAPI, React, Streamlit, Flask

Cloud & Platforms: AWS SageMaker, Azure (ML Studio, Blob Storage), Groq

Databases: PostgreSQL, MySQL, Oracle, SQL

DevOps & Tools: Docker, Nginx, Git, Bash, FAISS, Ollama, Whisper, pyannote, MSER, Power Automate

Spoken Languages: Arabic (Native), French (DELF B2), English (TOEIC)