# ABBAOUI **ACHRAF**

## Software Engineer | Machine Learning Engineer [NLP/CV]

in linkedin.com/in/achraf-abbaoui

github.com/abbaouiAchraf

@ achraf.abbaoui2001@gmail.com

M medium.com/achrafabbaoui

Ø abbaoui-achraf.netlify.app/

(+212)7 00 98 77 19

Rabat, Morocco



Innovative Computer Science Engineer specializing in AI and Machine Learning, with a proven track record in developing and deploying cutting-edge ML and AI solutions. With experience in developing and deploying advanced machine learning and AI based solutions. Being passionate about new technologies, I am seeking innovative projects where I can apply my skills and contribute to cutting-edge solutions.



#### **EDUCATION**

2019 - 2024

State Engineer National School of Applied Science - Tetouan

Field: Computer Science - Information System and Decision Support



## WORK EXPERIENCE

### Present July 2024

## Hiberus IT, Machine Learning Engineer, Full-time Employee

- Designed and implemented a scalable Retrieval-Augmented Generation (RAG) system, enhancing information retrieval by 30% and reducing search latency to sub-second levels for enterprise knowledge management in specialized domains (Industrial, Media, Government, etc.).
- > Fine-tuned Large Language Models (LLMs) (LLaMA, Phi, Mistral, BERT) for NLP tasks, achieving 95% accuracy while reducing training costs by 40% through low-bit quantization and efficient parameter pruning.
- > Developed an Autonomous Multi-Agent AI framework that automated customer support and data analysis workflows, cutting manual intervention by 50% and improving response times by 60%.
- Led the technical design and architecture of multiple AI projects, ensuring 100% on-time delivery while enhancing system scalability and performance.

Hugging Face	ace LLM LangChain LLamaInde		LLamaIndex	CrewAl Transformers		Pytorch	Vector Stores	OpenAl	Azure Cloud	
AWS FastAP	I Dock	ker 🛚	Pydantic	Pydantic-Al	Stream	nlit Gradio				

### July 2024 Jan 2024

### Hiberus IT, MACHINE LEARNING ENGINEER, Internship

- Designed and deployed a hybrid RAG system, integrating sparse and dense embeddings to improve context retrieval, reducing query resolution time by 35% and boosting user satisfaction.
- Developed AI proof-of-concepts (PoCs) for enterprise clients, surpassing baseline performance metrics by 25% and securing multiple new business contracts.

Hugging Face	LLM	LangChain	LLamaIndex	Pytorch	Vector Stores	OpenAl	Azure Cloud	AWS	FastAPI
Docker Stream	nlit								

## Aug 2023 July 2023

## MARSA MAROC Casablanca, DATA SCIENCE, Internship

- Developed a Vehicle License Plate Detection system with 97% accuracy, reducing manual verification workload by 60% and improving port security and logistics tracking.
- Created a Django/Angular real-time monitoring dashboard, streamlining data retrieval and enabling non-technical staff to generate reports 2x faster, improving operational decision-making.

OCR	Fine-tune	Deep Learning	Angular	Diango	II Al I	Object Detection

## Feb 2023

## Technocolabs Softwares Inc., DATA SCIENCE, Internship

#### Dec 2022

- Developed ML models predicting mortgage prepayment risk with 82% accuracy, enabling financial institutions to refine risk assessment and reduce default rates by 20%.
- Built interactive financial risk analysis dashboards, helping stakeholders optimize portfolio strategies and improve investment decision-making by 30%.

Scikit-learn   Pandas   NumPv   Machine Learning   Fla	ask    Docker    A	AWS
--	--------------------	-----

## Sep 2022

## Ministry of Energy Transition and Sustainable Development - Kingdom of Morocco, COMPUTER SCIENCE, Internship

## Aug 2022

- Developed a digital internship management platform, automating workflows and reducing processing time by 60%.
- Designed an optimized database and real-time dashboard, improving administrative efficiency in resource planning and reducing application processing delays by 42%.



#### SAAS PLATFORM FOR EXPENSE MANAGEMENT

**𝚱** Confidential source code

Developed a **full-stack SaaS platform** for **user expense management**, integrating Al-driven automation and intelligent financial insights. **Engineered a FastAPI backend** leveraging **Pydantic AI and autonomous agents** for intelligent expense tracking and categorization. Designed a **Next.js frontend** with a modern UI, supporting **OAuth authentication via Supabase** (Google, GitHub). Integrated **Stripe** for subscription management, enabling multiple pricing tiers and automated billing.

FastAPI Pydantic Al Autonomous Agents Next.js Supabase (OAuth) Stripe PostgreSQL

#### **IMAGE SEARCH ENGINE**

**©** Confidential source code

Developed an Al-powered image similarity search engine using EfficientNet for feature extraction, enabling scalable and accurate visual search. Implemented an optimized FAISS-based embedding storage, supporting real-time search over large image datasets. Built an interactive Streamlit-based frontend, allowing users to upload images and find similar content efficiently.

PyTorch EfficientNet FAISS Streamlit Pillow NumPy Docker Flask API

#### FINE-TUNING PHI3 MODEL FOR MEDICAL Q&A

₱ hf.co/AchrafABBAOUI/MedQuad\_Phi3\_QLoRA

Fine-tuned Microsoft Phi3 LLM for medical Q&A, optimizing domain-specific accuracy using the MEDQuaD dataset. Implemented QLoRA quantization to reduce memory usage and enable fine-tuning on resource-constrained environments. Deployed the model on Hugging Face Hub, making it accessible for research and production integration.

Microsoft Phi3 QLoRA MEDQuaD-dataset Generative Al Hugging Face

# **SKILLS**

Machine Learning & Al

Python, SQL, PL/SQL, PyTorch, Hugging Face, Transformers, LangChain, LlamaIndex, CrewAl, Fine-tuning LLMs, BERT, RAG, NLP, Computer Vision, Generative Al

Cloud & DevOps

Microsoft Azure, AWS, Docker, Kubernetes, MLflow, FastAPI, CI/CD, Terraform, Supa-

Software Development & Frameworks

FastAPI, Flask, Django, Angular, Next.js, Pandas, NumPy, Scikit-learn, OpenCV, Streamlit

Databases & Storage

MySQL, Po stgreSQL, Oracle Database, MongoDB, Vector Stores (FAISS, Pinecone, ChromaDB, Qdrant)

AI & Machine Learning Expertise

- ➤ End-to-end design, development, and deployment of Machine Learning and Deep Learning models
- Fine-tuning and optimization of Large Language Models (LLMs) for NLP, using techniques such as QLoRA and Low-bit Quantization

  Patricula Augmented Congretion (DAC) as store development for knowledge.
- ➤ Retrieval-Augmented Generation (RAG) system development for knowledgebased AI applications
- Advanced Natural Language Processing (NLP) tasks: Text Classification, Text Generation, Named Entity Recognition (NER)
- Computer Vision applications : Object Detection, Facial Recognition, Image Classification
- ➤ Building Autonomous AI Agents for multi-agent collaboration and decision-making
- Strong expertise in problem-solving and data-driven decision-making for AI applications

**Project Management & Software Engineering** 

Agile, Scrum, JIRA, Version Control (Git, GitHub Actions, GitLab CI/CD), Software Architecture, API Development

Other Tools & Concepts

Streamlit, OCR, MLOps, Model Serving (Triton, TensorRT), Model Distillation, Edge AI, AutoML, Stripe API (Subscription-based SaaS integration)

## A Z LANGUES

Français: Professional working proficiency (B2)
 Anglais: Professional working proficiency (C1)

> Arabe: Native proficiency



## **CERTIFICATIONS**