



Youssef Moutaouakkil

Data Scientist & MLOps Engineer

✉ youssefmoutaouakkil98@gmail.com  LinkedIn  Github

☎ +33782341408  Website & Portfolio




Profile

Data Scientist & MLOps Engineer with **2+ years of experience** specializing in **enterprise AI solutions** and **production ML systems**. Proven track record in developing and deploying **end-to-end data pipelines** and **scalable ML architectures** for leading management consulting firms including **BCG** and **PwC**. Experience spans **MLOps infrastructure**, **multimodal document processing**, **RAG-based knowledge systems**, and **custom LLM fine-tuning**, delivering solutions that transformed manual workflows and drove significant business value across diverse industry use cases.

Professional Experience

Data Scientist

2022 – present

Infomineo 

Founding member of the Data Analytics department, spearheaded end-to-end AI and data science initiatives for initial clients including Fortune 500, delivering solutions from proof-of-concept to production:

- Architected a **multimodal document processing system** leveraging **Vision-Language models** combined with specialized **OCR pipelines** to extract and normalize financial data from 10,000+ heterogeneous documents in multiple languages and formats, implementing an **active learning framework** with user feedback loop that reduced a 6-month manual process to 1 month, deployed as a scalable microservice architecture using **FastAPI** backend, **React** frontend, and **GCP Kubernetes Engine** with **Cloud Storage** for document management and **Cloud Functions** for on-demand processing
- Implemented **DPO fine-tuning** on **OS LLM** with industry-specific datasets, leveraging **HF** and **W&B**, transforming a 2-month manual data processing workflow into a 2-week automated solution with integrated quality checks.
- Developed an automated internal market intelligence platform implementing **NLP pipelines** for entity extraction and **sentiment analysis**, utilizing **Playwright** for robust web scraping, with **Airflow**-orchestrated **ETL pipelines** deployed on **GCP**, featuring real-time analytics dashboards that auto-update via **cron jobs** and advanced semantic search functionality using **RAG architecture**, **VectorDB** and **Google's Grounding techniques**, enabling daily processing and analysis of 100GB+ unstructured market data.

Data Scientist - Intern

2022

Mohammed VI Polytechnic University 

Designed and implemented an **end-to-end Computer Vision based mobile application** (From Model training to Mobile application) that significantly increased the collection and quality of snow depth data.

Education

Masters. Industrial Engineering, Data Science Focus

ESITH

Specialized in Process Optimization, ML Systems Design, Industrial IoT Data Integration & engineering, Pipeline Architecture...

CPGE - Prepa Math-Physics

- Intensive undergraduate program for competitive entrance exams into top French/Moroccan engineering schools.

Courses in mathematics, physics, computer science, Algorithms, Advanced Calculus, Topology, Thermodynamics...

Languages

English

Bilingual

French

Bilingual

Arabic

Native

Skills

Tools :

- Python, C++, Bash
- Scikit-learn, PyTorch, SQL, Git, Docker, Kubernetes, FastAPI, Spark
- Cloud (Mostly GCP & AWS)

Technical knowledge:

- Deep learning : NLP, LLMs & Computer Vision
- Machine Learning (Supervised & Unsupervised)
- A/B testing & Statistical Analysis
- MLOps & CI/CD

Certificates

- Tensorflow Developer [🔗](#)

- Deep Learning and Neural networks [🔗](#)

Projects

ModelMe AI: Professional Imagery via Fine-Tuning [🔗](#)

Professional photo AI generator where you upload your own images and a custom model is trained on it

Ask your PDF: Streamlit-integrated RAG Assistant [🔗](#)

Streamlit-based RAG for Document Q&A: Upload PDFs, Automatically embed content into a vector database, and use your OpenAI key for AI-powered search and answers.