Mouad Elbaz

Junior Data Engineer | Python, SQL, ETL/ELT

Agadir, Morocco | +212 751 81 12 01 | elbaz.ouad1249@gmail.com | github.com/moad-cod | Portfolio

Education

Universite Sultan Moulay Slimane 2026
Bachelor in Big Data *Morocco*

ISTA Agadir 2023 – 2025
Technicien Specialise en Reseaux et Systemes *Agadir, Morocco*

Lycee Qualifiant 2022 – 2023
Baccalaureat Sciences Physiques *Morocco*

Professional Experience

Data Engineering & Machine Learning Intern (PFE) Apr. 2026 – Jul. 2026
Prefecture d'Agadir Ida-Outanane *Agadir, Morocco*

- Designed an asynchronous **ELT pipeline** integrating **200,000+ records** from GLPI ITSM, OCS Inventory, and security logs using **Medallion Architecture**.
- Orchestrated data workflows with **Apache Airflow** and modeled **Bronze, Silver, and Gold** analytical layers using **dbt**.
- Developed a multilingual retrieval pipeline combining **BM25**, dense vector search, Reciprocal Rank Fusion, Cross-Encoder re-ranking, and **Qdrant**.
- Integrated **XGBoost** ticket classification and Isolation Forest anomaly detection into **FastAPI** services and a **Next.js** KPI dashboard.

Developer Intern Apr. 2025 – Jun. 2025
Prefecture d'Agadir Ida-Outanane *Agadir, Morocco*

- Developed an enterprise attendance management application with role-based access control (**RBAC**), authentication, session management, IP compliance analysis, and audit logging.

Projects

CCSR – Data Ingestion & AI Research Platform | *Python, Airflow, Celery, PostgreSQL, Docker* GitHub

- Developed a local-first data and AI research platform with reproducible ingestion workflows, experiment tracking, and data lineage.
- Implemented **Bronze/Silver/Gold data pipelines** using **MinIO**, PostgreSQL, configurable chunking, and deterministic indexing into Qdrant.
- Integrated **Apache Airflow and Celery** for workflow orchestration, Redis for caching/events, and FastAPI for data services, with tests for retries and failure recovery.

Qwen MedQuAD – Data Processing Pipeline | *Python, Pandas, Data Validation, QLoRA* GitHub

- Built a reproducible dataset preparation pipeline for **16,357 cleaned medical QA records**, including cleaning, normalization, validation, and leakage auditing.
- Organized **14,344 normalized question groups** using group-safe train/validation/test splitting to support reliable downstream ML experiments.

PlantGuard AI – ML Data Pipeline | *Python, PyTorch, Data Processing, Docker* GitHub

- Prepared and processed **70,000+ images** across 38 plant disease categories for reproducible CNN model evaluation.
- Benchmarked CNN, ResNet, and EfficientNet-B3 and deployed the resulting classification application using Docker and Streamlit.

Technical Skills

Programming: Python, SQL, Java
Data Engineering: ETL/ELT, Apache Airflow, dbt, PySpark, Apache Spark, Hadoop, Medallion Architecture, Data Pipelines
Databases & Storage: PostgreSQL, MySQL, Redis, MinIO, Qdrant, FAISS
Backend & Integration: FastAPI, REST APIs, Celery, Data Processing, Pandas, NumPy
DevOps & Cloud: Docker, Git, Linux, Oracle Cloud Infrastructure (OCI)
Machine Learning: Scikit-learn, XGBoost, PyTorch, RAG, Hybrid Retrieval

Certifications & Languages

Certifications: Introduction to Data Warehousing (365 Data Science), Machine Learning Specialization (DeepLearning.AI), Network Security (Cisco), Microsoft Learn Developer & Cybersecurity Pathways
Languages: Arabic – Native | French – Intermediate | English – Advanced