

KHADIJA OUBAHA

khadijaoubaha111@gmail.com | Casablanca | LinkedIn: khadija oubaha

AI & Computer Science Engineer | Data Science · Machine Learning · Data Analysis · Data Engineering

Profile

Recently graduated AI & Computer Science Engineer with three technical internships in machine learning research, NLP/LLM systems, and enterprise AutoML tooling, skilled in building end-to-end data pipelines, deploying ML models, and turning raw data into decision-ready insights.

Work Experience

Machine Learning Research Intern

UM6P, Benguerir, Morocco

Feb 2026 – Present

- Designed, trained, and benchmarked 3 deep learning models on 10,000 physics simulations, raising displacement-field accuracy from 87.7% to 92.2% via physics-informed loss.
- Built an iterative prediction model that generalized to resolutions up to 8x finer than training, raising unseen-resolution accuracy from 0–36% to 60–92% using 5x less training data.
- Ran cluster-scale training and data generation, delivering predictions 10x faster than the original physics solver.

AI Development Intern

WELEARN, Casablanca, Morocco

Jun 2025 – Jul 2025

- Built a Python/Pandas ETL pipeline feeding an 8-KPI Streamlit/Plotly dashboard, giving WELEARN's pedagogical team self-serve visibility across 3 training programs.
- Engineered a hybrid RAG pipeline (FAISS + BM25) using MiniLM embeddings to index course PDFs, resolving semantic-search failures on technical queries and validating against a 60-question test suite.
- Deployed Meta-Llama-3-8B-Instruct as a CUDA-enabled Docker service on Hugging Face Spaces, secured behind a token-authenticated proxy.

Machine Learning Intern

Oracle, Casablanca, Morocco

Jul 2024 – Sep 2024

- Integrated explainability for forecasting datasets in Oracle's AutoMLx, engineering regression tests to benchmark runtime and memory against Scikit-learn, and resolving edge cases in feature-less univariate data.
- Added 2 regression-test stages to Oracle's Jenkins CI/CD, automating per-dataset pass/fail logging and error-traceback capture pre-release.
- Standardized an 18-metric recommendation results schema and unified column naming across 2 explainer functions, replacing internal class names with human readable labels, to eliminate reporting inconsistencies for end users.

Education

Engineering Degree in Artificial Intelligence and Computer Science

ENSAM Casablanca, Morocco

Sep 2021 – Jun 2026

Projects

SecureDoc AI – Intelligent Document Anonymization Platform

[GitHub]

Java · Spring Boot · Docker · Kubernetes · AWS EC2 · RabbitMQ · DJL/BERT-NER · GitLab CI/CD · Prometheus/Grafana

- Built a 4-service microservices platform (Spring Boot/Java 17) using RabbitMQ for async processing and PostgreSQL/MongoDB/MinIO for storage.
- Built a hybrid PII-detection engine combining regex with a BERT-NER model served via Deep Java Library, then redacted sensitive regions directly on the source PDF with PDFBox.
- Deployed the full stack on Kubernetes via AWS EC2, with a self-hosted GitLab CI/CD runner and Prometheus/Grafana monitoring.

Medicinal Plant Identifier – Offline Computer Vision App

[GitHub]

TensorFlow/Keras · InceptionResNetV2 · Transfer Learning · TensorFlow Lite · Flutter · SQLite

- Fine-tuned InceptionResNetV2 on a cleaned, augmented image pipeline to classify 80 medicinal plant species, reaching 87.37% test accuracy over 50 epochs with early stopping.
- Converted the model to TensorFlow Lite for fully offline, on-device inference, and built the Flutter app (Android/iOS) with real-time camera capture, instant identification, and toxicity alerts.
- Designed a hybrid data layer pairing a JSON knowledge base (names, uses, toxicity, remedies) with a 3-table SQLite schema for users, history, and favorites — fully functional with zero network connection.

Skills

Programming Languages: Python, SQL, Java

Libraries & Frameworks: Pandas, NumPy, Scikit-learn, TensorFlow/Keras, PyTorch, LangChain, Spring Boot

Tools & Platforms: Power BI, Streamlit, Plotly, AWS (EC2, Route 53), Docker, Kubernetes, PostgreSQL, MongoDB, Hugging Face

Dev & Collaboration: Git, GitLab CI/CD, Jenkins, RabbitMQ, Jira, Confluence, Linux

Certificates: Machine learning with Python (IBM), AWS Certified Cloud Practitioner, Power BI (365 Data Science)

Languages: English, French, Arabic, Spanish