

Idriss Khattabi

E-mail: drisskhattabi6@gmail.com | Phone: +212 6 35 24 57 97

Portfolio: drisskhattabi6.github.io/id-kh

LinkedIn: linkedin.com/in/idriss-khattabi-b3a266235/

GitHub: github.com/drisskhattabi6



About me

Data Engineer with a Master's in AI & Data Science, skilled in designing, building, and maintaining robust data pipelines, data warehousing, and ETL processes. Proficient in Python, SQL, and cloud/data engineering tools, with experience in data cleaning, transformation, and integration from multiple sources. Familiar with big data frameworks, workflow automation, and scalable architectures. Analytical, detail-oriented, and passionate about enabling data-driven decision-making through high-quality, reliable data infrastructure.

Education

M.S.T in Artificial Intelligence and Data Science

Faculty of Sciences and Technologies of Tangier

Oct 2023 – July 2025

Tangier, Morocco

BS in Computer Science and Mathematics

Faculty multidisciplinary of Larache

Oct 2019 – July 2023

Larache, Morocco

Experience

AI / Data Engineer Intern, TNA Consulting – Casablanca

Feb 2025 – July 2025

- Designed and developed an AI SaaS platform to support financial institutions in interpreting and leveraging official central bank publications & documents across Africa.
- Engineered an automated and scalable **Data pipeline** for scraping & collecting documents from websites, cleaning, transforming & chunking documents, and loading into a Vector database.
- Built an intelligent **chatbot** using RAG Technique capable of answering user questions on African Central Bank documents.
- Used Tools : **Python, Marker, Streamlit, Apach Airflow, ETL Pipeline, Flask, Ollama, Faiss, RAG, LLM, Selenium, Docker**

Research Assistant, Faculty of Poly-disciplinary – Larache

March 2023 – June 2023

- Designed and implemented an optimization system to improve the consistency of pairwise comparison matrices in the **AHP**, leveraging **GA** and **PSO**, with Developing a Django web application that makes it easier to generate AHP matrices.
- Co-authored **three** research papers with **Dr. Chakir Tajani**, papers examples :
 - A Metaheuristic Approach to Improve Consistency of the Pairwise Matrix in AHP
 - Generating a Set of Consistent Pairwise Comparison Test Matrices in AHP using PSO

Soft Skills

Analytical thinking and problem-solving – Fast learning – Teamwork and collaboration – Agility and Adaptability – Curious for new technologies

Technical Skills

- Programming & Scripting:** Python, SQL, Java, JavaScript, C, Bash, Solidity
- Data Analysis & Visualization:** NumPy, Pandas, Matplotlib, Seaborn, Plotly, Power BI, Power Query
- Big Data & Distributed Systems:** Apache Spark (PySpark), Apache Kafka, Hadoop, Apache Airflow
- Machine & Deep Learning & NLP:** Scikit-learn, TensorFlow, Keras, PyTorch, XGBoost, MLflow, NLTK, SpaCy
- Generative AI & LLMs:** RAG, Prompt Engineering, Fine-tuning, LangChain, LangGraph, LlamaIndex, n8n
- Databases:** MySQL, SQL Server, PostgreSQL, MongoDB, Neo4j, IPFS, FAISS, Chroma, Qdrant, Milvus
- Web Development:** HTML, CSS, Tailwind CSS, Javascript, ReactJS/NextJs, Django, Flask, FastAPI, Streamlit
- DevOps & Tools:** Linux, Git/GitHub, Docker, Microsoft Azure, Selenium, BeautifulSoup, PyTest

Languages

Arabic: Native proficiency | **English:** Professional (B2) | **French:** Professional (B2)

Certifications

Data Manipulation – DataCamp

Data Analyst with Python – DataCamp

Associate Data Engineer – DataCamp

Microsoft Azure Fundamentals (AZ-900) – DataCamp

Projects

Real-time Twitter Post Sentiment Analysis

[GitHub Repo]

- Created a real-time system to predict sentiment from Twitter posts by streaming posts using Kafka and predicting in real-time PySpark using Machine Learning, and visualizing data in real time.
- Tools Used: PySpark, Kafka, Docker, Django, MongoDB, ChartJS

Build and Analyze E-Commerce Data Warehouse

[GitHub Repo]

- Implement an ETL pipeline to create Data Warehouse for E-commerce product and supplier analytics with dashboards and reports.
- Tools Used: Python, Pandas, SQLAlchemy, MySQL, Power BI

Data Space for Electronic Medical Records

[GitHub Repo]

- Comprehensive platform for managing, processing, and exploring healthcare data.
- Tools Used: Python, MySQL, ETL Pipeline, RESTful API

Fake News Detection

[GitHub Repo]

- Fine-tune a pre-trained BERT model to detect fake news articles using a labeled dataset. Includes data pre-processing, model training, evaluation, and deployment via Flask web app.
- Tools Used: Python, Flask, BERT, Fine-tuning

Data Analysis & Machine Learning Desktop Application

[GitHub Repo]

- Python desktop app built with CustomTkinter for analyzing qualitative and quantitative datasets and applying multiple machine learning algorithms via an intuitive GUI.
- Tools Used: Python, Tkinter / CustomTkinter, Scikit-learn, Pandas, Matplotlib

Stock Market Analysis and Prediction

[GitHub Repo]

- Implements a Deep Learning model to analyze and predict stock market prices using historical stock data.
- Tools Used: Python, PyTorch, Jupyter Notebook

YouTube Transcript Summarizer Bot

[GitHub Repo]

- AI-powered workflow to extract transcripts from YouTube videos via Telegram bot and summarize them in .txt files.
- Tools Used: n8n, Telegram Bot API, RapidAPI, JavaScript, Google Gemini

AI Scraper

[GitHub Repo]

- Interactive tool to scrape, process, and structure web data using local or cloud-based LLMs.
- Tools Used: Python, Selenium, Crawl4AI, Streamlit, Ollama, Sambanova, Pandas

Age, Gender & Emotion Detection using CNN

[GitHub Repo]

- Detecting person's age, gender, and emotional state from facial images using CNN.
- Tools Used: Python, OpenCV, Tkinter, TensorFlow/Keras, CNN, NumPy, Pretrained Models

Moroccan Meal Recipes Chatbot

[GitHub Repo]

- AI chatbot using RAG to provide authentic Moroccan recipes with ingredients, steps, and optional images; compatible with local or cloud LLMs.
- Tools Used: Python, Streamlit, LangChain, Ollama, Sambanova, Faiss DB