

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



À propos de moi

Data Engineer titulaire d'un Master en Intelligence Artificielle et Science des Données, compétent dans la conception, la construction et la maintenance de pipelines de données robustes, de data warehouses et de processus ETL. Maîtrise de Python, SQL et des outils de cloud/data engineering, avec une expérience en nettoyage, transformation et intégration de données issues de multiples sources. Esprit analytique et passionné par la prise de décision basée sur les données grâce à des infrastructures de données fiables et de haute qualité.

Éducation

Master S.T. en Intelligence Artificielle et Data Science Oct. 2023 – Juil. 2025
Faculté des Sciences et Techniques de Tanger Tanger, Maroc

Licence en Informatique et Mathématiques Oct. 2019 – Juil. 2023
Faculté Pluridisciplinaire de Larache Larache, Maroc

Expérience

Stagiaire Ingénieur IA / Data, TNA Consulting – Casablanca Fév. 2025 – Juil. 2025

- Conception et développement d'une plateforme SaaS basée sur l'IA pour aider les institutions financières à interpréter et exploiter les publications et documents officiels des banques centrales à travers l'Afrique.
- Mise en place d'un **pipeline de données** automatisé et évolutif pour le scraping et la collecte de documents depuis des sites web, leur nettoyage, transformation, segmentation, puis intégration dans une base de données vectorielle.
- Développement d'un **chatbot intelligent** utilisant la technique RAG, capable de répondre aux questions des utilisateurs sur les documents des banques centrales africaines.
- Outils utilisés : **Python, Marker, Streamlit, Apache Airflow, Pipeline ETL, Flask, Ollama, Faiss, RAG, LLM, Selenium, Docker**

Assistant de recherche, Faculté Pluridisciplinaire – Larache Mars 2023 – Juin 2023

- Conception et implémentation d'un système d'optimisation visant à améliorer la cohérence des matrices de comparaison par paires dans la méthode **AHP**, en utilisant les algorithmes **GA** et **PSO**, avec le développement d'une application web Django facilitant la génération de matrices AHP.
- Co-auteur de **trois** articles de recherche avec **Dr. Chakir Tajani**, exemples :
 - 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

Compétences personnelles

Esprit analytique et résolution de problèmes – Apprentissage rapide – Travail d'équipe et collaboration – Agilité et adaptabilité – Curiosité pour les nouvelles technologies

Compétences techniques

- **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
- **Bases de données** : MySQL, SQL Server, PostgreSQL, MongoDB, Neo4j, IPFS, FAISS, Chroma, Qdrant, Milvus
- **Développement Web** : HTML, CSS, Tailwind CSS, JavaScript, ReactJS/NextJS, Django, Flask, FastAPI, Streamlit
- **DevOps & Outils** : Linux, Git/GitHub, Docker, Microsoft Azure, Selenium, PyTest

Langues

Arabe : Langue maternelle | **Anglais** : Professionnel (B2) | **Français** : Professionnel (B2)

Certifications

Associate Data Analyst in SQL – DataCamp
Data Manipulation – DataCamp

Data Analyst with Python – DataCamp
Associate Data Engineer – DataCamp

Projets

Analyse en temps réel du sentiment des publications Twitter

[GitHub Repo]

- Conception d'un système temps réel pour prédire le sentiment des publications Twitter.
- Outils utilisés : PySpark, Kafka, Docker, Django, MongoDB, ChartJS

Construction et analyse d'un Data Warehouse pour l'e-commerce

[GitHub Repo]

- Mise en œuvre d'un pipeline ETL pour créer un Data Warehouse dédié à l'analyse des produits et fournisseurs d'un site e-commerce, avec création de tableaux de bord et rapports.
- Outils utilisés : Python, Pandas, SQLAlchemy, MySQL, Power BI

Data Space pour les dossiers médicaux électroniques

[GitHub Repo]

- Plateforme complète pour la gestion, le traitement et l'exploration de données médicales.
- Outils utilisés : Python, MySQL, ETL Pipeline, RESTful API

Détection de Fake News

[GitHub Repo]

- Fine-tuning d'un modèle pré-entraîné BERT pour détecter les articles de fausses nouvelles à partir d'un jeu de données étiqueté, incluant prétraitement, entraînement, évaluation et déploiement via une application web Flask.
- Outils utilisés : Python, Flask, BERT, Fine-tuning

Application de bureau pour l'analyse de données et le Machine Learning

[GitHub Repo]

- Application de bureau Python développée avec CustomTkinter permettant d'analyser des jeux de données qualitatifs et quantitatifs, et d'appliquer plusieurs algorithmes de Machine Learning via une interface graphique intuitive.
- Outils utilisés : Python, Tkinter / CustomTkinter, Scikit-learn, Pandas, Matplotlib

Analyse et prédiction du marché boursier

[GitHub Repo]

- Mise en œuvre d'un modèle de Deep Learning pour analyser et prédire les prix boursiers à partir de données historiques.
- Outils utilisés : Python, PyTorch, Jupyter Notebook

YouTube Transcript Summarizer Bot

[GitHub Repo]

- Workflow d'automatisation intelligent permettant d'extraire et de résumer les transcriptions de vidéos YouTube envoyées via un bot Telegram, livrées en fichier .txt.
- Outils utilisés : n8n, Telegram Bot API, RapidAPI, JavaScript, Google Gemini

AI Scraper

[GitHub Repo]

- Outil interactif pour extraire, traiter et structurer des données web à l'aide de LLMs locaux ou hébergés dans le cloud.
- Outils utilisés : Python, Selenium, Crawl4AI, Streamlit, Ollama, Sambanova, Pandas

Détection de l'âge, du genre et des émotions via CNN

[GitHub Repo]

- Détection de l'âge, du genre et de l'état émotionnel d'une personne à partir d'images faciales en utilisant un réseau de neurones convolutif (CNN).
- Outils utilisés : Python, OpenCV, Tkinter, TensorFlow/Keras, CNN, NumPy, Pretrained Models

Chatbot de recettes marocaines

[GitHub Repo]

- Chatbot IA utilisant RAG pour fournir des recettes marocaines authentiques avec ingrédients, étapes et images optionnelles, compatible avec des LLMs locaux ou cloud.
- Outils utilisés : Python, Streamlit, LangChain, Ollama, Sambanova, Faiss DB