

MOTASSIM Hamza

Data Engineer Student

✉ motassimhamza99@gmail.com ☎ +212 6 78 33 19 93 🌐 www.hamzaae.xyz 📺 Hamza MOTASSIM

Data engineering student at the National School of Applied Sciences Al-Hoceima, dedicated to transforming raw information into actionable insights, analyst, ambitious, innovative, and with a strong ability to acquire new knowledge. Looking for an End-of-studies internship in the field of data.

Experience

- **Data Engineer Intern at Shiftbricks - Remote** *July - September 2024*
 - Developed an AI assistant capable of interacting with documents and responding to queries using Retrieval-Augmented Generation (RAG). This project involved building a full-stack application using Python, FastAPI, LangChain, mongoDB, and Next.js, among other technologies like Docker.
- **Data Engineer Intern at OCP Group - On site** *July 2024*
 - Engineered a data pipeline to process instrument measurements, classify values as good or bad, and predict future values. The project utilized Python libraries (Pandas, Matplotlib, Scikit-learn, TensorFlow, Keras), integrated with Kafka, and deployed on Azure (VM, Data Factory, Blob Storage, SQL DB, Web App).
- **Software Engineer Intern at ENSAH IT Departement - Hybrid** *June - July 2023*
 - Developed an application for managing new student admissions and processing their applications by multiple user roles. The project was implemented using PHP, MySQL, HTM/CSS/JS.
- **Data Engineer Intern at OCP Group - On site** *August 2023*
 - Assisted in the digitalization of the water desalination zone by transforming data through pipelines. This project involved using Spring, React, Azure, Kafka PostgreSQL and other technologies.

Education

- **Lycée Ihsan** *2020*
 - High School degree in Physics
- **National School of Applied Sciences of Al-Hoceima** *2020 – present (to be completed in 2025)*
 - Data engineering field

Projects

- IDFS** *github.com/hamzaae/IDIFS-master*
 - Built a distributed file system for fast file storage and retrieval. It includes a master node ("FATHER MASTER") for metadata and multiple slave nodes ("CHILDS") for data. The system handles thousands of nodes, supports parallel and distributed computing, and uses RPYC for communication between components, ensuring fault tolerance with data replication.
- Darija Comments Sentiment Analysis** *lfahim.tech*
 - Involves leveraging various ML and DL algorithms and fine-tuning the BART model with Darija. This enables businesses to categorize comments as either positive or negative and scraping Hesperess articles to determine Sentiments percentage. The deployment of this refined model is facilitated through MLOps.
- ZeroRatt** *zero-ratt.vercel.app*
 - Created a website that helps students prepare for exams by allowing them to upload a file, which then generates quizzes to test their readiness. The project utilizes LangChain, Retrieval-Augmented Generation (RAG), Next.js, and other technologies.
- ThisIsNLP** *thisisnlp.vercel.app*
 - Developed a website using Next.js to share structured courses on Natural Language Processing (NLP). The content is based on the knowledge gained from studies, certifications, and internships.

Real-Time People Count Pipeline

github.com/hamzaae/Real_Time_People_Monitor

- Lambda architecture pipeline for counting people in different zones of a market video and generating insights from using yolov5 with Python, big data tools (kafka, hdfs, spark, airflow), Power BI visualization and Docker.

Data Pipelines

github.com/hamzaae

- Different data engineering pipelines employing multiple sources, transformations, and loading mechanisms ([Github Trending Repos](#) / [Avito Product Track](#) / ...). Utilizing a variety of technologies, including sql/nosql dbs, Big Data Tools (Kafka, Spark ...), and Cloud, to seamlessly handle data processing tasks.

Patent Analysis

github.com/hamzaae/Big-Data-Project-Patent-Analysis

- E-Patent is an academic comprehensive project aimed at analyzing patents related to Artificial Intelligence (AI) in education. It involves data collection (APIs and Webscraping), storage (MongoDb, PostgreSQL, HDFS), transformation (Spark) and visualization (Flask, PowerBI) using various big data technologies and frameworks. The primary goal is to provide personalized and general analyses of patents through a user-friendly web application.

Xou-Dou-Qui (Jungle Game)

github.com/hamzaae/XouDouQi

- Academic project using Java Swing with an implementation of Minimax algorithm to play against with different levels and using sockets to play against friends.

Activities

- Ensah Coding Challenge Winner, Tech Experience October 2023
- Participation in the 1337 Think AI Hackathon May 2023
- Participation in the Moroccan National Programming Contest February 2024
- Datai Club President at Ensah 2023 - 2024
- OFFICE AUTOMATION SOFTWARE & PROGRAMMING - Social center of proximity 2021
- Coursera & DataCamp +21 [Certifications](#) for Data & Ai Engineer 2022 - current

Skills

1. Hard Skills

Programming languages & Frameworks :

- Python, Java (Spring), JS (react & next) and R

DataBases :

- SQL (PostgreSql, MySql, SQL Server, OracleDB) - NoSql (MongoDb, Cassandra) - PL/SQL

Big Data :

- Hadoop Ecosystem (HDFS - Spark - Hive - Airflow - Kfaka)
- Linux, Shell, Bash scripting
- Data Cleaning / Visualization

Artificial Intelligence :

- Machine Learning / Deep Learning / LLMs

DevOPS :

- Git & GitHub , Docker, Azure Cloud

Conception & Modelisation :

- Agile / Scrum - UML

2. Soft Skills

Languages :

- Arabic (Native)
- English (Proficient)
- French (Proficient)

Skills :

- Problem solving, Critical thinking, Teamwork, Adaptability