# Hamza **EL BELGHITI**

First-year master's student in Data Science and Engineering, experienced with several data science tools & techniques such as nlp, sentiment analysis, and working with llm's like langchain to develop chatbots and RESTful api's. Worked on data engineering projects that involved building ETL pipelines and using cloud technologies like AWS and GCP.


Hamagistral.me | Hamagistral | Hamza El Belghiti | hamza.lbelghiti@gmail.com | +212-660081086


## EDUCATION

### MS. DATA SCIENCE AND ENGINEERING | 2022-2024

University Mohammed V, Faculty of Sciences of Rabat

**Relevant Coursework:** Data Science with Python, Business Intelligence, Datamining, DSA, Java

### BASc. APPLIED MATHEMATICS | 2019-2022

University Hassan II, Faculty of Sciences and Techniques of Mohammedia

**Relevant Coursework:** SQL, Databases, Calculus, Linear Algebra, Probability, Statistics (R), ML

## PROJECTS

### ANALYZING DATA ENGINEERS JOB LISTINGS ON GLASSDOOR | 2023

DATA SCIENCE & DATA ENGINEERING PROJECT | GITHUB DEMO

- Scraped over 2500 data engineers job listings from Glassdoor US using Selenium in a period of 6 weeks and stored the raw data in an AWS S3 bucket.
- Built a data pipeline that runs weekly using Mage, it extracts data from the raw S3 bucket, transforms it, merges it with previous weeks data, and finally loads it into a cleaned S3 bucket.
- Analyzed the job listings (EDA) and predicted data engineers salaries in the USA using a Random Forest Regressor algorithm, and deployed it on the web via Streamlit.

**Technologies :** Python, Selenium, AWS S3, Boto3, Mage, Seaborn, Scikit-learn, Streamlit

### ANALYZING NYC TAXI TRIP RECORDS | 2023

DATA ENGINEERING PROJECT | GITHUB DEMO

- Modeled NYC Taxi Trip Records data into star schema (dimensions and facts tables).
- Desgined a data pipeline with MageAI running on Google Compute Engine that extracts data from Google Storage and stores it in a data warehouse using BigQuery.
- Visualized data findings using an interactive dashboard on Looker Studio.

**Technologies :** Python, Google Cloud Storage, Compute Instance, Mage, BigQuery, Looker

### CUSTOMER SERVICE CHATBOT FOR DECATHLON MOROCCO | 2023

LLM AND SENTIMENT ANALYSIS PROJECT | GITHUB DEMO

- Scraped reviews of popular Decathlon products using Beautiful Soup, cleaned the raw data, explored the data, and then used NLTK and Vader to assess the sentiment of the review descriptions.
- Extracted the company's docs like the return policies and warranty details, from the website, splitted the docs and then embedded them using the OpenAI Ada model and stored them in Pinecone.
- Developed a chatbot using the gpt-3.5-model that uses the Pinecone index to provide relevant answers to customers questions using the company's documents, deployed using Streamlit.

**Technologies :** Python, BeautifulSoup, NTLK, Vader, Langchain, OpenAI, Pinecone, Streamlit

### BRANDGENIE : AI BRANDING ASSISTANT | 2023

WEB & AI PROJECT | GITHUB DEMO

- Created a web application deployed on Vercel that assists e-commerce entrepreneurs in finding brand names, slogans, keywords, and creative advertising copy.
- Developed an API using FastAPI that generates branding elements using the GPT3 model from OpenAI, hosted on AWS Lambda, and integrated with API Gateway.
- Integrated the API in the front-end created with Nextjs and used Firebase for authentication.

**Technologies :** Python, FastAPI, OpenAI API, AWS (Lambda + API Gateway), Docker, Firebase Auth, Typescript, React/Nextjs, Tailwindcss, Vercel

### GPTUBE : CHATGPT FOR YOUTUBE VIDEOS | 2023

STREAMLIT STUDENT CHALLENGE x LLM PROJECT | GITHUB DEMO

- Developed a Streamlit app that summarizes YouTube videos and generates responses to questions based on transcripts generated using OpenAI Whisper and Langchain.
- Reduced usage costs from $0.03 per minute of video to $0.006 by implementing the gpt-3.5-turbo model and using a vector database locally.

**Technologies :** Python, Whisper, Langchain, ChromaDB, Streamlit

## SKILLS

### PROGRAMMING

Python • Java • C • R

### DATABASES

Oracle • SQLite • MySQL • PostgreSQL • MongoDB • Snowflake • Pinecone

### DATA & MACHINE LEARNING

Pandas • Numpy • Seaborn • Matplotlib Scikit-learn • TensorFlow • Keras • NLTK Mage • Airflow • PySpark • Selenium • Kafka • PowerBI • Looker Studio

### WEB DEVELOPMENT

HTML/CSS • Java/Typescript • MERN • Nextjs • Django • FastAPI • Streamlit

### CLOUD

AWS • GCP • Firebase

### OTHER TOOLS

Linux • Docker • Git • Zenhub (Agile)

### LANGUES

Arabic • English • French

## CERTIFICATES

### DATA SCIENTIST WITH PYTHON

Issued by **DataCamp**

December 2022

### NEURAL NETWORKS AND DEEP LEARNING

Issued by **DeepLearning.AI**

November 2022

### MACHINE LEARNING WITH PYTHON

Issued by **FreeCodeCamp**

April 2022

### CS50 WEB PROGRAMMING

Issued by **HARVARD**

February 2022

## EXTRACURRICULAR

### STREAMLIT FOR EDUCATION

Student Ambassador | 2023

### FSTM IT CLUB

Member | 2021-2022

## INTERESTS

- Blogging
- Chess