Mehdi BEN ATTIA

Data Scientist

✉ mehdi.benattia01@gmail.com   📞 +33 6 49 67 86 02

📍 Paris, France   in Mehdi Ben Attia

○ Git Repository

Professional Experiences

Machine Learning Engineer - Apprenticeship   09/2024 – 09/2025
Safran
- Designed and trained RNN/LSTM models for engine health monitoring using sensor data using Python, PyTorch, GitLab, and AWS.
- Developed a RAG (Retrieval-Augmented Generation) system to improve internal document search using LiteLLM, LlamaIndex, Qdrant, HyDE, LangChain, and Claude.

AI Engineer - Internship   01/2024 – 07/2024
Scaylab   Paris, France
- Fine-tuned large language models (LLMs) to build recommendation and expert-matching systems, using Python, LlamaIndex, Mistral, Claude, LangChain, and ChromaDB.
- Deployed AI systems in containerized environments using Docker, and orchestrated deployments via AWS ECR and ECS for scalable performance.

AI Developer - Freelance   09/2023 – 01/2024
Symmetryk   Paris, France
- Built a RAG system to optimize document retrieval using Python, LlamaIndex, Faiss, Qdrant, LangChain, and multiple indexing techniques (TF-IDF, BM25, embeddings).
- Collected and analyzed user feedback to improve search accuracy using reinforcement learning from human feedback (RLHF).

Data Scientist - Internship   06/2023 – 08/2023
CarNet   Paris, France
- Designed and implemented a recommendation system in Python to guide users toward optimal repair centers, using TF-IDF, KNN, and cosine similarity.
- Deployed the system to production via AWS stack: S3, SageMaker, Lambda, and ECS, ensuring scalability and reliability.

Achievements

1st place - AI4Health Hackathon ⧉   05/2023
- Fine-tuned a large language model (GPT-2) to identify and classify the severity of side effects from medication usage.
- Built a demo interface using Streamlit; utilized HuggingFace Transformers.

1st place - OpenGovData Hackathon ⧉   04/2023
- Developed an intelligent solution to simplify and automate access to mammography services using public health datasets.

1st place - CardioHackathon ⧉   03/2023
- Designed a smart decision-support tool to assist physicians in treating patients with heart failure using predictive modeling and medical datasets.

Education

MS Expert Big Data Engineer
Université de Technologie Troyes
2024 – 2025

Prépa intégrée / Software Engineer - Data Science
Ecole Supérieure Privée d'Ingénierie et de Technologies ⧉
2019 – 2024

Personal Projects

Skill-Based Curriculum Platform
- Python, Web Scraping, Data Preprocessing, NLP, Recommandation System, SkillNER, Data Visualization, Power BI.

Chronic Kidney Disease Prediction
- Business Understanding, Data Preprocessing, training and fine-tuning a machine learning model, evaluating, deployment Flask.

Skills

AI / Data Science
Python, Machine Learning, Deep Learning, NLP, LLM, RAG, TensorFlow, Pytorch, Data Analysis, Power BI, Dataiku, LlamaIndex, Ollama, HuggingFace, Langchain.

Databases
SQL, MongoDB, Qdrant, ChromaDB, ElasticSearch, Neo4J.

DevOps / MLOps
Github, GitLab, AWS, Docker, Airflow, Spark.

Certifications

- Microsoft Certified - Power BI Data Analyst Associate ⧉
- Microsoft Certified - AI Fundamentals
- PMI - Project Mangament Ready ⧉

Languages

- Arabe / Francais / Anglais - Courant