



Mehdi BEN ATTIA

Data Scientist

✉ mehdi.benattia01@gmail.com

☎ +33 6 49 67 86 02

📍 Paris, France

🌐 Mehdi Ben Attia

📁 Git Repository

Experiences Professionnelles.

Ingénieur Machine Learning - Alternance

09/2024 – 09/2025

Safran

- Concevoir et entrainer des réseaux de neurones RNN/LSTM pour l'estimation de la santé des moteurs à partir de données capteurs, en utilisant Python, PyTorch, GitLab et AWS.
- Développer un RAG pour améliorer la recherche documentaire interne en exploitant LiteLLM, LlamaIndex, Qdrant, HyDE, LangChain, et Claude.

Ingénieur IA - Stage

01/2024 – 07/2024

Scaylab

- Fine-tuning de LLMs pour développer des systèmes de recommandation et de matching entre startups et experts, en utilisant Python, LlamaIndex, Mistral, Claude, LangChain et ChromaDB.
- Déployer des modèles sur AWS via Docker, avec intégration complète sur ECR et ECS pour garantir une exécution fiable et scalable des solutions IA.

Développeur IA - Freelance

09/2023 – 01/2024

Symmetryk

- Développer un système RAG pour optimiser la recherche documentaire à l'aide de Python, LlamaIndex, Faiss, Qdrant, LangChain, et des techniques d'indexation comme TF-IDF, BM25 et les embeddings.
- Collecter et exploiter les retours utilisateurs pour améliorer la pertinence des résultats via un apprentissage par renforcement (RLHF) avec Python.

Data Scientist - Stage

06/2023 – 08/2023

CarNet

- Implémenter un système de recommandation en Python pour orienter les utilisateurs vers les centres de réparation optimaux, basé sur TF-IDF, KNN et la similarité cosinus.
- Déployer la solution sur AWS (S3, SageMaker, Lambda, ECS), garantissant un service évolutif et disponible en production.

Accomplissements

1ère place en AI4Health Hackathon

05/2023

Fine-tuning d'un LLM pour identifier la gravité des effets secondaires apparaissant après la prise de certains médicaments: StreamLit, Bert, Transformers.

1ère place en OpenGovData Hackathon

04/2023

Piloter une solution intelligente visant à simplifier et automatiser l'accès aux mammographies.

1ère place en CardioHackathon

03/2023

Conception d'un outil intelligent pour aider les médecins à traiter leurs patients souffrant d'insuffisance cardiaque.

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

Projets Réalisés

Plateforme de curriculum basée sur les compétences.

- Python, Web Scraping, Data Preprocessing, NLP, Recommandation System, SkillNER, Data Visualization, Power BI.

Prédiction de la maladie rénale chronique.

- 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.

Gestion des Bases de données

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

DevOps / MLOps

Github, GitLab, AWS, Docker, Airflow, Spark.

Certificats

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

Languages

- Arabe / Francais / Anglais - Courant