



AYOUB EL HOUDRI Data Scientist

✉ ayoubelhoudri@gmail.com

☎ +33 7 83 22 67 03

📍 Paris, France

🔗 y-aoub.github.io

🌐 linkedin.com/in/elhoudriay

🐙 github.com/y-aoub

PROFIL

Data Scientist disposant d'une expérience en conception, développement et mise en œuvre de solutions d'intelligence artificielle à grande échelle. Compétences en machine learning, deep learning et IA générative, couvrant l'ensemble du cycle de vie des projets : recherche et prototypage, déploiement en production, suivi et amélioration continue.

EXPÉRIENCE PROFESSIONNELLE

AI Engineer (CDI), Waalaxy

09/2024 – à ce jour | Montpellier, France

- Création d'un système multiagents pour générer des messages LinkedIn personnalisés, augmentant le taux de réponse de 35 %.
- Développement d'un moteur de recherche vectoriel pour la prospection, basé sur 700 millions de profils LinkedIn.
- Amélioration de la pertinence des résultats du moteur de recherche de 30 % grâce à un fine-tuning sur un vocabulaire métier.
- Mise en place d'outils de surveillance pour assurer la performance et la fiabilité des modèles en production.

Outils : Python (LangGraph, Unsloth), Langfuse, MongoDB (Atlas Vector Database), GCP (Vertex AI, BigQuery)

Data Scientist & Data Engineer (Indépendant), AEH

11/2022 – 09/2024 | Paris, France

- Conception et mise en œuvre de modèles de machine learning et de workflows d'IA générative pour des clients de divers secteurs.
- Architecture et développement de flux ETL, de l'extraction automatisée au maintien à jour du data warehouse.
- Déploiement de modèles de machine learning en production pour services à fort trafic, suivant les bonnes pratiques MLOps.

Outils : Python (LangChain, PySpark, Scikit-learn, MLflow), Snowflake, Pinecone, GCP (Cloud Run, Vertex AI)

Data Scientist (Stage), Median Technologies

05/2023 – 10/2023 | Nice, France

- Implémentation d'un modèle GAN pour la segmentation 3D des poumons sur scans CT, atteignant 96 % de précision.
- Optimisation d'un pipeline de détection de tumeurs, en réduisant les faux négatifs de 30 %.
- Conception d'un système d'aide à la classification des tumeurs selon leur type, validé par deux radiologues.

Outils : Python (PyTorch, MONAI, OpenCV), PostgreSQL, Apptainer (ex-Singularity), 3DSlicer

Data Scientist (Stage), Karmen

06/2022 – 09/2022 | Paris, France

- Conception d'un système d'extraction des données à partir de documents financiers, réduisant la saisie manuelle de 75 %.
- Implémentation et déploiement d'un système de détection de fraude sur des données tabulaires avec 82 % de précision.
- Intégration de pratiques MLOps pour superviser les modèles en production et détecter les dérives sur les données d'entrée.

Outils : Python (PyTorch, PyTesseract, OpenCV, XGBoost), Snowflake, Grafana, Dataiku

Machine Learning Engineer (Stage), Digimind Labs

05/2021 – 08/2021 | Berlin, Allemagne

- Développement d'un générateur de données synthétiques 3D d'entraînement.
- Entraînement du modèle Pixel2Mesh pour la reconstruction 3D à partir d'une seule vue.
- Déploiement du modèle pour une utilisation sur les chaînes de production.

Outils : Python (Tensorflow, OpenCV), Blender API, AWS (Amazon S3, AWS SageMaker)

FORMATION

Master Recherche en Informatique - Intelligence Artificielle et Systèmes

2022 – 2023 | Cergy, France

Complexes, Université CY Cergy Paris / ENSEA / Laboratoire ETIS (CNRS)

Diplôme d'Ingénieur en Mathématiques Appliquées et Informatique - Data

2020 – 2023 | Cergy, France

Science, CY Tech

COMPÉTENCES TECHNIQUES

Langages de programmation: Python, R, SQL, Bash, C

Machine Learning et IA: PyTorch, Scikit-learn, XGBoost, Hugging Face Transformers, LangChain, LangGraph

MLOps et versioning: MLflow, Langfuse, Docker, Git, GitHub Actions

Traitement des données: Pandas, Polars, NumPy, SciPy, PySpark, OpenCV

Infrastructure et déploiement: Google Cloud Platform, Kubernetes, FastAPI, Flask

Bases de données: Snowflake, MySQL, PostgreSQL, MongoDB, Pinecone

LANGUES

Français : Bilingue

Anglais : Courant

Espagnol : Débutant