



DHIA HMILA

Data Scientist Senior chez AXA France

✉ dhia.hmila@dasquare.xyz ☎ +33610201606 📍 92700, Colombes, France
🌐 hmladhia.github.io 🌐 hmladhia 🔗 dhia-hmila 📄 Dhia Hmila



EXPÉRIENCE PROFESSIONNELLE

Data Scientist Senior

Équipe Data Science & IA, AXA France

📅 Juin 2021 - Aujourd'hui 📍 Nanterre, France

- Développé un système RAG basé sur un LLM pour répondre aux questions des agents d'assurance et réduire la charge du service support.
- Animé des ateliers avec des ingénieurs ML, des ingénieurs de données et des data scientists pour développer le Guide des Meilleures Pratiques MLOps pour AXA France.
- Créé un template Cookiecutter pour de nouveaux projets ML afin d'accélérer le time-to-market (TTM) chez AXA.
- Construit des systèmes d'IA pour l'extraction d'informations structurées à partir de documents (CNI, permis, etc.), permettant le traitement de plus de 6 millions de documents avec PySpark.
- Conçu des packages de document processing réutilisables (dont certains en open-source) pour accélérer le développement de moteurs de document processing.
- Automatisé l'indexation de 70 % des documents entrants d'AXA France en utilisant des systèmes d'IA réentraînaibles pour la classification de documents.
- Réduction de la charge de travail du service d'assistance en développant un système RAG basé sur un LLM pour répondre aux questions des agents d'assurance.
- Construit des pipelines Azure ML réutilisables pour l'entraînement continu (CT), facilitant le déploiement des modèles dans le Model Registry.
- Conçu des templates Azure DevOps de CI/CD, améliorant la qualité du code et le temps de livraison des projets.
- Coordonné le processus d'annotation avec les métiers pour garantir des ensembles de données de haute qualité à travers divers projets.
- Dirigé la COP Python d'AXA en organisant des présentations mensuelles sur des sujets tels que les tests, le web scraping et la qualité du code.
- Animé des sessions de formation pour la manipulation de données avec Python pour plus de 30 collaborateurs d'AXA, renforçant ainsi leurs compétences

Graduate Program - Data Scientist

Équipe Data, Fraud, Waste & Abuse, AXA France

📅 Mai 2020 - Juin 2021 📍 Nanterre, France

- Construit un outil de détection de fraudes et d'abus piloté par l'IA, permettant des économies dépassant 150 000 € en 2022.
- Conçu et entraîné plusieurs modèles de détection de fraude, améliorant significativement les performances de la détection.

Data Scientist

Shift Technology

📅 Avril 2019 - Mars 2020 📍 Paris, France

- Appliqué des techniques de NLP pour cluster-iser les notes des gestionnaires de sinistres, révélant des patterns de fraude émergents grâce à une visualisation avancée.

A PROPOS DE MOI

Je suis un Data Scientist Senior spécialisé dans la détection de fraudes/anomalies et la classification de documents. J'ai travaillé sur une panoplie de sujets ML dont des sujets Computer Vision ou encore NLP/NLU. Je possède également une solide expérience autour des pratiques de MLOps, pour garantir une intégration fluide des modèles dans les environnements de production.

CURSUS

Double diplôme d'ingénieur

🎓 ENSTA ParisTech

🎓 ENIT

📅 2016 - 2019

📍 France - Tunisia

Spécialisé en Intelligence Artificielle

Cycle préparatoire

🎓 IPEIT

🎓 Esprit Prépa

📅 2014 - 2015

📍 France - Tunisia

Spécialisé en Math & Physique Rang 10 au concours d'entrée aux écoles d'ingénieurs en Tunisie parmi plus de 3000 candidats.

COMPÉTENCES

Programmation

🐍 Python

Rust

📄 JavaScript

C/C++

Intelligence Artificielle

scikit-learn

TensorFlow

PyTorch

Transformers

Generative AI

MLOps

MLflow

🔄 Github Actions

🏗️ Azure Pipelines

🐳 Docker

Kubernetes

Natural Language Processing

- Développé et déployé de nouveaux modèles d'IA pour la solution de détection de fraude de Shift, améliorant ainsi les offres aux clients.

Stage de recherche

U2IS - ENSTA Paris

📅 Mai 2018 - Août 2018

📍 Palaiseau, France

- Créé un ensemble de données d'images synthétiques en simulant un agent dans un environnement interne, en utilisant le ray tracing pour un rendu d'image réaliste.
- Exploré l'oubli catastrophique dans l'apprentissage incrémental des composantes principales des images générées.

PROJETS

📦 mlflow: Outil d'orchestration pour les projets MLflow.

mlflow orchestration experiment-tracking open-source

📅 Novembre 2023 - Aujourd'hui

Un outil d'orchestration qui améliore les projets MLflow en fournissant une gestion intégrée des workflows multi-étapes.

- Utilisé des graphes orientés acyclique (DAG) pour gérer des workflows ML complexes, garantissant une exécution logique des tâches.
- Implémenté un tri topologique pour optimiser l'ordre des tâches en fonction des dépendances, améliorant ainsi l'efficacité des workflows.
- Amélioré les projets MLflow avec une gestion claire des dépendances pour des expériences en plusieurs étapes.
- Intégré des outils de visualisation pour une compréhension intuitive des workflows et des dépendances.

📦 pysira: publication de CVs dans différents formats.

Github Actions Jinja2 Templates jsonschema

📅 Février 2023 - Aujourd'hui

Un outil pour exporter des fichiers 'jsonresume' vers divers formats (HTML, TeX, PDF) et langues.

- Développé un pipeline CI/CD utilisant GitHub Actions pour automatiser la publication de CV dans plusieurs formats et thèmes.
- Permis l'exportation de fichiers 'jsonresume' vers HTML, TeX, PDF et d'autres formats.
- Démontré la fonctionnalité avec un CV en direct : hmladhia.github.io, hmladhia.github.io/cv.pdf.

📦 Splitter: Outil de Traitement de Documents

Traitement de Documents RAG open-source

📅 Décembre 2021 - Aujourd'hui

Un package open-source offrant des outils pour traiter divers types de documents (PDF, TIFF, etc.) en images et extraire du texte en utilisant Python.

- Prend en charge plusieurs formats de documents pour un traitement polyvalent.
- Architecture personnalisable avec support de plugins pour des fonctionnalités étendues.

📄 Spot-Language: Prédiction des langage de programmation

spaCy

Langchain

Pydantic-ai

RAG

LLMs

ReACT

Data Wrangling

pandas

Polars

PySpark

SQL

Computer Vision

OpenCV

Vision Transformers

YOLO

Cloud

Azure ML

Databricks

LANGUES

Arabe
Anglais
Français



CERTIFICATIONS



Machine Learning Engineering for Production (MLOps) Specialization

DeepLearning.AI

📅 2022 2022



Natural Language Processing Specialization

DeepLearning.AI

📅 2024 2024



Cloud-Native Development with OpenShift and Kubernetes Specialization

Red Hat

📅 2023 2023



Rust Fundamentals

Duke University

📅 2024 2024



Rust for DevOps

Duke University

📅 2024 2024



Rust for LLMops (Large Language Model Operations)

Duke University

📅 2024 2024



DevOps, DataOps, MLOps

Duke University

📅 2024 2024



MLOps Tools: MLflow and Hugging Face

Duke University

📅 2023 2023



Introduction to Large Language Models

Google Cloud

📅 2023 2023



TOEIC Listening and Reading Test (Score 985/990)

ETS Global

📅 2018 2018

[Classification](#)[Flask](#)[NLP](#)[LIME](#) Janvier 2020

Un modèle de classification pour détecter les langages de programmation dans des extraits de code.

- Construit un jeu de données d'entraînement provenant de repositories github publiques.
- Construit une expérience ML pour entraîner un modèle de classification pour la détection de langages de programmation.
- Amélioré l'interprétabilité du modèle en utilisant LIME pour de meilleures insights sur les prédictions.
- Déployé une application web de démonstration utilisant Flask pour présenter le modèle ML.