

Dhia Hmila



+33610201606 | dhia.hmila@dasquare.xyz | [hmiladhia.github.io](https://github.com/hmiladhia) |

[hmiladhia](#) | [dhia-hmila](#) | [Dhia Hmila](#) | [dhia-hmila](#)

Je suis un Data Scientist Senior spécialisé dans la détection de fraudes/anomalies

Expérience Professionnelle

Équipe Data Science & IA, AXA France

DATA SCIENTIST SENIOR

Nanterre, France

2021 - Aujourd'hui

MLOps OCR NLP LLM RAG

- 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

Équipe Data, Fraud, Waste & Abuse, AXA France

GRADUATE PROGRAM - DATA SCIENTIST

Nanterre, France

2020 - 2021

Python Machine Learning Fraud Detection Anomaly Detection scikit-learn

- 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.

Shift Technology

DATA SCIENTIST

Paris, France

2019 - 2020

Python C# Fraud Detection NLP Clustering

- Appliqué des techniques de NLP pour cluster-iser les notes des questionnaires de sinistres, révélant des patterns de fraude émergents grâce à une visualisation avancée.
- Développé et déployé de nouveaux modèles d'IA pour la solution de détection de fraude de Shift, améliorant ainsi les offres aux clients.

Compétences

Langues Arabe ♦ Anglais ♦ Français

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 spaCy ♦ Langchain ♦ Pydantic-ai ♦ RAG ♦ LLMs ♦ ReACT

Data Wrangling pandas ♦ Polars ♦ PySpark ♦ SQL
Computer Vision OpenCV ♦ Vision Transformers ♦ YOLO
Cloud Azure ML ♦ Databricks

Projets

mlflower: Outil d'orchestration pour les projets MLflow.

[HTTPS://GITHUB.COM/HMILADHIA/MLFLOWER](https://github.com/hmiladhia/mlflower)

2023 - Aujourd'hui

mlflow orchestration experiment-tracking open-source

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.

[HTTPS://GITHUB.COM/HMILADHIA/PYSIRA](https://github.com/hmiladhia/pysira)

2023 - Aujourd'hui

Github Actions Jinja2 Templates jsonschema

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.
- Démonstré la fonctionnalité avec un CV en direct : hmiladhia.github.io, hmiladhia.github.io/cv.pdf.

Splitter: Outil de Traitement de Documents

[HTTPS://GITHUB.COM/AXAFRANCE/AXA-FR-SPLITTER](https://github.com/AXAFrance/axa-fr-splitter)

2021 - Aujourd'hui

Traitement de Documents RAG open-source

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

[HTTPS://SPOT-LANGUAGE.ONRENDER.COM/](https://spot-language.onrender.com/)

2020 - 2020

Classification Flask NLP LIME

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.

Cursus

ENSTA ParisTech / ENIT

DOUBLE DIPLÔME D'INGÉNIEUR - INTELLIGENCE ARTIFICIELLE

2016 - 2019

IPEIT / Esprit Prépa

CYCLE PRÉPARATOIRE - MATH & PHYSIQUE

2014 - 2015

- Rang 10 au concours d'entrée aux écoles d'ingénieurs en Tunisie parmi plus de 3000 candidats.

Certifications

- 2022 Machine Learning Engineering for Production (MLOps) Specialization, [DeepLearning.AI](#)
- 2024 Natural Language Processing Specialization, [DeepLearning.AI](#)
- 2023 Cloud-Native Development with OpenShift and Kubernetes Specialization, [Red Hat](#)
- 2024 Rust Fundamentals, [Duke University](#)
- 2024 Rust for DevOps, [Duke University](#)
- 2024 Rust for LLMOps (Large Language Model Operations), [Duke University](#)
- 2024 DevOps, DataOps, MLOps, [Duke University](#)
- 2023 MLOps Tools: MLflow and Hugging Face, [Duke University](#)

- 2023 **Introduction to Large Language Models**, [Google Cloud](#)
- 2018 **TOEIC Listening and Reading Test (Score 985/990)**, [ETS Global](#)