

Dhia Hmila

☎ +33610201606 | ✉ dhia.hmila@dasquare.xyz | ✉ hmladhia.github.io |

📷 hmladhia | 🔗 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

- Développement de systèmes d'IA d'extraction d'information basé pour lire des documents (permis de conduire, carte d'identité nationale, etc.). Les modèles construits ont permis de traiter plus de 6 millions de documents.
- Création d'une gamme de packages Python réutilisables pour le traitement de documents (OCR, détection d'objets, etc.). Les packages sont conçus pour expérimenter rapidement avec différents modèles (Strategy Pattern).
- Automatisation de l'indexation de 70% des documents entrants d'AXA France (flux sinistre) en développant un système d'IA de classification de documents (exposé en tant qu'API).
- 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.
- Élaboration de pipelines Azure ML réutilisables (CT - Continuous Training) pour réentraîner automatiquement les systèmes d'IA et les publier dans le Registre des modèles.
- Conception de modèles de pipelines CI/CD Azure DevOps à utiliser dans les projets de l'équipe.
- Supervision du processus de recrutement, intégration des nouveaux membres de l'équipe et facilitation du transfert efficace des connaissances.
- Coordination du processus d'annotation avec les annotateurs et les experts métiers pour garantir des datasets cohérents et de qualité.
- Direction de l'initiative "Python COP" d'AXA, organisant une série d'événements mensuels sur la programmation portant sur des sujets tels que les tests, le web scraping, le packaging, la qualité du code, etc.
- Animation d'une série de formations (sur Python, la manipulation des données avec pandas & pyspark, les bonnes pratiques de code) pour plus de 30 collaborateurs d'AXA.

Équipe Data, Fraud, Waste & Abuse, AXA France

GRADUATE PROGRAM - DATA SCIENTIST

Nanterre, France

2020 - 2021

- Conception et développement d'un outil de détection avancée des fraudes, gaspillage et abus piloté par l'IA, agissant à différentes échelles (sinistres, prestataires de santé, bénéficiaires). Cet outil novateur a joué un rôle clé dans la réalisation d'économies dépassant les 150 000 euros en 2022.
- Conception et entraînement de divers modèles de détection de fraudes pour repérer les patterns frauduleux, dans le cadre de l'outil de détection des FWA (Fraud, Waste & Abuse).

Shift Technology

DATA SCIENTIST

Paris, France

2019 - 2020

- Mis en œuvre des techniques de Traitement du Langage Naturel (NLP) pour regrouper les notes des gestionnaires de sinistres, en utilisant des méthodes de visualisation avancées afin de déceler et d'identifier les patterns de fraude émergents.
- Élaboration de nouveaux modèles d'IA pour la solution de détection de fraude de Shift et déploiement de la solution pour un nouveau client.

U2IS - ENSTA Paris

STAGE DE RECHERCHE

Palaiseau, France

2018 - 2018

- Simulation d'un agent dans un environnement artificiel et création d'un dataset d'images où on peut percevoir l'effet de l'oubli catastrophique en utilisant le Raytracing pour rendre des images réalistes..
- Évaluation de la méthode "Elastic Parameters Consolidation" proposée par un article de recherche pour remédier à l'oubli catastrophique qui surgit lors de l'apprentissage incrémental.

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.

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

Outil d'orchestration de jobs d'entraineemnt basé sur MLflow Projects.

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

Outil en ligne de commande permettant d'exporter des CVs au format 'jsonresume' vers différents formats (HTML, TeX, PDF, etc.) et langues (Anglais, Français, etc.).

- Création d'un pipeline CI/CD (Actions Github) pour publier des CV dans différents thèmes, formats et langues.
- Ce CV a été créé en utilisant pysira : hmladhia.github.io, hmladhia.github.io/cv.pdf, hmladhia.github.io/fr/kendall ou hmladhia.github.io/resume.pdf

nbmanips: Découper, fusionner et convertir vos Notebooks

[HTTPS://PYPI.ORG/PROJECT/NBMANIPS/](https://pypi.org/project/nbmanips/)

2021 - Aujourd'hui

Jupyter Databricks Zeppelin

Un package open-source d'outils de manipulation de Notebooks via des scripts python ou une CLI.

Spot-Language: Prédiction du langage de programmation

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

2020 - 2020

Classification Flask NLP Compression de Modèles

Spot-Language est un modèle de classification des langages utilisés dans des sources de code.

- Construction de dataset d'entrainement provenant de repositories github publiques.
- Déploiement d'une application Web pour servir de démo au modèle IA.

Dmail: Envoi d'e-mails au format Markdown

[HTTPS://PYPI.ORG/PROJECT/DMAIL/](https://pypi.org/project/dmail/)

2020 - 2020

email markdown downloads>100k

Un package python qui permet d'envoyer des e-mails au format markdown

Compétences

| | |
|---------------------------|--|
| Langues | Arabe Anglais Français |
| Intelligence Artificielle | Anomaly Detection Tensorflow TFX scikit-learn NLP nltk spaCy Gensim pyod |
| Data Wrangling | pandas pyspark SQL ETL |
| Langues de programmation | Python Rust JavaScript C/C++ C# |
| MLOps | Github Actions Azure Pipelines Git Linux Docker Azure ML Databricks |
| Web Scraping | selenium requests BeautifulSoup4 |

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](#)