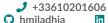


DHIA HMILA

Senior Data Scientist at AXA France

- @ dhiahmila@gmail.com
- hmiladhia.github.io
- dhia-hmila





92700, Colombes, France M Dhia Hmila



PROFESSIONAL EXPERIENCE

Senior Data Scientist

Data Science Team, AXA France

- June 2021 Present
- Nanterre. France
- Automated the indexing of 70% of AXA's incoming documents by developing a retrainable document classification AI System (exposed as an API).
- Developed text extraction AI Systems using computer vision to read documents (Driver's license, National ID card, etc.). The models processed over 6M documents in batches using PvSpark Jobs
- Developed a range of reusable python packages for document processing (OCR, object detection, etc.) The packages are designed using the strategy pattern to quickly experiment with different algorithms / models.
- Created a YAML-based templating system to streamline the creation of Azure ML pipelines.
- Built reusable Azure ML Pipelines (Continuous Training) to automatically retrain Al Systems and publish to Model Registry.
- Designed Azure DevOps CI/CD/CT pipeline templates to be used in the Team's projects.
- Led the hiring process, onboarding new team members and facilitating efficient knowledge transfer.
- Coordinated the annotation process with annotators and SMEs (subject matter experts) to ensure consistent datasets across various projects.
- Led AXA's 'Python COP' initiative, organizing a series of monthly programming talks and events on topics such as testing, web scraping, packaging, code quality, etc.
- Teached a series of training sessions/courses on Python, Data Manipulation (pandas & pyspark). Software Engineering best practices for over 30 AXA collaborators.

Graduate Data Scientist

Data, Fraud, Waste & Abuse Team, AXA France

- May 2020 June 2021
- Nanterre, France
- Engineered an advanced Al-driven Fraud, Waste, and Abuse Detection Tool with a wide-ranging impact across multiple scopes, including Claims, Health Providers, Beneficiaries, and more. This innovative tool played a pivotal role in achieving savings exceeding 150k euros in 2022.
- Designed and trained various fraud detection models for fraudulent patterns as part of the FWA detection tool.

Data Scientist

Shift Technology

- April 2019 March 2020
- Paris, France
- Applied Natural Language Processing (NLP) techniques to cluster Claim managers' notes, leveraging advanced visualization methods to uncover and identify emerging fraud patterns.
- Built new AI Models for Shift's fraud detection solution & Deployed their solution for a new customer.

ABOUT

I am a Senior Data Scientist and I specialize in fraud/anomaly detection and document classification. My expertise extends to leading-edge technologies, including machine learning, computer vision, and natural language processing.

EDUCATION

Engineering double degree

ENSTA ParisTech

2016 - 2019

France - Tunisia

Specialized in Artificial Intelligence

Preparatory School

IPEIT

🙎 Esprit Prépa

2014 - 2015

France - Tunisia

Specialized in Mathematics & Physics Ranked 10th in the entrance exam for engineering schools among 3,000 candidates.

SKILLS

Machine Learning

mlflow scikit-learn Tensorflow pytorch transformers trax pyod huggingface pytorch-lightning

Computer Vision

OpenCV **Vision Transformers** torchvision VGG16 YOLO Feature Matching tesseract paddleOCR easy-ocr

Natural Language Processing

langchain nltk spaCy Gensim torchtext

Data Wrangling

pandas pyspark **SQL**

Research Internship

U2IS - ENSTA Paris

- **May 2018 August 2018**
- Palaiseau, France
- Created a synthetic image dataset by building a simulation of an agent in an in-house environment and using Raytracing to render realistic images.
- Studied Catastrophic Forgetting in Incremental Learning of Principal Components of generated Images

PROJECTS

mlflower: Lightweight orchestration tool for mlflow projects

mlflow orchestration experiment-tracking

November 2023 - Present

Tool to extend mlflow projects with inbuilt multi step workflow orchestration

pysira: publish resumes in different formats

Github Actions | Jinja2 Templates | jsonschema

February 2023 - Present

CLI tool to export 'jsonresume' files to different formats (HTML, TeX, PDF, etc.) and languages.

- Created a CI/CD pipeline (Github Actions) to publish resumes in different themes, formats and languages
- This resume was created using pysira: hmiladhia.github.io, hmiladhia.github.io/cv.pdf, English hmiladhia.github.io/fr/kendall or hmiladhia.github.io/resume.pdf
 French

nbmanips: Split, merge and convert IPython Notebooks

Jupyter Databricks Zeppelin

May 2021 - Present

An open-source package containing a collection of tools to manipulate Notebooks via a python or CLI.

- Developed a python package/CLI to manipulate (Split, Merge, ...) Jupyter Notebooks
- Used CI/CD piplines (Github Actions) to Lint, Test and publish the package to Pypi

☐ Spot-Language: Programming Language Detection

Classification Flask NLP Lime

January 2020

Spot-Language is a classification model to detect the language used in code snippets.

- Assembled a training dataset using public repositories on GitHub.
- Built an ML experiment to train an NLP model for language classification.
- Improved model interpretability using Lime.
- Deployed a demo Web Application of the ML model using flask.

Programming Languages

Python Rust JavaScript C/C++
C#

MLOps

Github Workflows Azure Pipelines

Git Linux Docker Kubernetes

Cloud

Azure ML Databricks Heroku

Web Scraping

selenium requests BeautifulSoup4

Web

HTML CSS Js JavaScript fastapi

A flask Dash

LANGUAGES

Arabic English French German



CERTIFICATES

Machine Learning Engineering for Production (MLOps)

DeepLearning.AI = 2022 2022

Natural Language Processing
DeepLearning.Al 2024 2024

Cloud-Native Development with Open-Shift and Kubernetes Specialization

DeepLearning.AI = 2023 2023

Rust Fundamentals
Duke University

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