

# Dhia Hmila

+33610201606 | dhia.hmila@dasquare.xyz | hmiladhia.github.io | hmiladhia | dhia-hmila | Dhia Hmila | dhia-hmila



Senior Data Scientist with expertise in fraud detection, document classification, and MLOps.

## Professional experience

### Data Science Team, AXA France

SENIOR DATA SCIENTIST

Nanterre, France

2021 - Present

MLOps OCR NLP LLM RAG

- Developed an LLM-based RAG system, reducing help-desk workload by providing instant answers to insurance agents.
- Led workshops with ML Engineers, Data Engineers, and Data Scientists to develop the MLOps Best Practices Guide for AXA.
- Created a Cookiecutter template for new ML projects to accelerate time-to-market (TTM) at AXA France.
- Developed text extraction AI systems using computer vision, processing over 6M documents in batches with PySpark, reducing processing time by 30%.
- Created reusable Python packages for document processing, enabling rapid experimentation with various algorithms.
- Automated the indexing of 70% of incoming documents through a retrainable document classification AI system, improving efficiency.
- Built reusable Azure ML pipelines for continuous training, ensuring models are always up-to-date in the Model Registry.
- Designed CI/CD/CT pipeline templates in Azure DevOps, improving project delivery timelines.
- Led the hiring process and onboarding of new team members, facilitating efficient knowledge transfer and team integration.
- Coordinated the annotation process with subject matter experts to ensure high-quality datasets across various projects.
- Led AXA's 'Python COP' initiative, organizing monthly talks on topics such as testing, web scraping, and code quality.
- Conducted training sessions on Python and data manipulation for over 30 AXA collaborators, enhancing their skills.

### Data, Fraud, Waste & Abuse Team, AXA France

GRADUATE DATA SCIENTIST

Nanterre, France

2020 - 2021

Python Machine Learning Fraud Detection Anomaly Detection scikit-learn

- Engineered an AI-driven Fraud, Waste, and Abuse Detection Tool, achieving savings exceeding 150k euros in 2022.
- Designed and trained multiple fraud detection models, significantly improving detection accuracy.

### Shift Technology

DATA SCIENTIST

Paris, France

2019 - 2020

Python C# Fraud Detection NLP Clustering

- Applied NLP techniques to cluster claim managers' notes, revealing emerging fraud patterns through advanced visualization.
- Developed and deployed new AI models for Shift's fraud detection solution, enhancing client offerings.

## Skills

#### Languages

Arabic ♦ English ♦ French ♦ German

#### Programming

Python ♦ Rust ♦ JavaScript ♦ C/C++

#### Machine Learning

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

#### Cloud Technologies

Azure ML ♦ Databricks

## Education

### ENSTA ParisTech / ENIT

ENGINEERING DOUBLE DEGREE - ARTIFICIAL INTELLIGENCE

2016 - 2019

### IPEIT / Esprit Prépa

PREPARATORY SCHOOL - MATHEMATICS & PHYSICS

2014 - 2015

- Ranked 10th in the entrance exam for engineering schools among 3,000 candidates.

# Projects

---

## mlflow: Lightweight orchestration for mlflow projects

<https://github.com/hmiladhia/mlflow>

2023 - Present

mlflow orchestration experiment-tracking DAG open-source

An orchestration tool that enhances MLflow projects by providing built-in multi-step workflow management.

- Utilized Directed Acyclic Graphs (DAGs) for managing complex ML workflows, ensuring logical task execution.
- Implemented topological sorting to optimize task order based on dependencies, enhancing workflow efficiency.
- Enhanced MLflow projects with clear dependency management for multi-step experiments.
- Integrated visualization tools for intuitive understanding of workflows and dependencies.

## pysira: Publish resumes in different formats

<https://github.com/hmiladhia/pysira>

2023 - Present

GitHub Actions Jinja2 Templates jsonschema open-source

A CLI tool for exporting 'jsonresume' files to various formats (HTML, TeX, PDF) and languages.

- Developed a CI/CD pipeline using GitHub Actions to automate resume publishing in multiple formats and themes.
- Enabled export of 'jsonresume' files to HTML, TeX, PDF, and other formats.
- Demonstrated functionality with a live resume: [hmiladhia.github.io](https://hmiladhia.github.io), [hmiladhia.github.io/cv.pdf](https://hmiladhia.github.io/cv.pdf).

## splitter: Document Processing Tool

<https://github.com/AXAFrance/axa-fr-splitter>

2021 - Present

Document Processing RAG open-source

An open-source package that provides tools for processing various document types (PDF, TIFF, etc.) into images and extracting text using Python.

- Supports multiple document formats for versatile processing.
- Customizable architecture with plugin support for extended functionality.

## nbmanips: Split, merge and convert IPython Notebooks

<https://pypi.org/project/nbmanips/>

2021 - Present

Jupyter Databricks Zeppelin open-source

An open-source toolset for manipulating Jupyter Notebooks via Python or CLI.

- Created a Python package and CLI for splitting, merging, and converting Jupyter Notebooks.
- Set up CI/CD pipelines using GitHub Actions for linting, testing, and publishing to PyPI.
- Contributed to the open-source community by providing tools for Jupyter Notebook management.

## Spot-Language: Programming Language Detection

<https://spot-language.onrender.com/>

2020 - Present

Classification Flask NLP LIME

A classification model for detecting programming languages in code snippets.

- Assembled a training dataset from public GitHub repositories on Github
- Built an ML experiment to train a classification model for programming language detection.
- Enhanced model interpretability using LIME for better insights into predictions.
- Deployed a demo web app using Flask to showcase the ML model.

# Certificates

---

2022	Machine Learning Engineering for Production (MLOps) Specialization, <a href="#">DeepLearning.AI</a>
2024	Natural Language Processing Specialization, <a href="#">DeepLearning.AI</a>
2023	Cloud-Native Development with OpenShift and Kubernetes Specialization, <a href="#">Red Hat</a>
2024	Rust Fundamentals, <a href="#">Duke University</a>
2024	Rust for DevOps, <a href="#">Duke University</a>
2024	Rust for LLMops (Large Language Model Operations), <a href="#">Duke University</a>
2024	DevOps, DataOps, MLOps, <a href="#">Duke University</a>
2023	MLOps Tools: MLflow and Hugging Face, <a href="#">Duke University</a>
2023	Introduction to Large Language Models, <a href="#">Google Cloud</a>
2018	TOEIC Listening and Reading Test (Score 985/990), <a href="#">ETS Global</a>