DHIA HMILA

Senior Data Scientist at AXA France

@ dhia.hmila@dasquare.xyz   📞 +33610201606   📍 92700, Colombes, France
🌐 hmiladhia.github.io   ⌂ hmiladhia   in dhia-hmila   Ⓜ Dhia Hmila
📖 dhia-hmila

PROFESSIONAL EXPERIENCE

Senior Data Scientist
Data Science Team, AXA France

📅 June 2021 - Present   📍 Nanterre, France

- 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.

Graduate Data Scientist
Data, Fraud, Waste & Abuse Team, AXA France

📅 May 2020 - June 2021   📍 Nanterre, France

- 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.

Data Scientist
Shift Technology

📅 April 2019 - March 2020   📍 Paris, France

- 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.

ABOUT

Senior Data Scientist with expertise in fraud detection, document classification, and MLOps. Proven track record in developing AI systems that process millions of documents and streamline operations. Strong background in machine learning, computer vision, and natural language processing, with a focus on delivering impactful solutions.

EDUCATION

Engineering double degree
🎓 ENSTA ParisTech   🎓 ENIT

📅 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

Programming

🐍 Python   Rust   JS 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

PROJECTS

🧊 mlflower: Lightweight orchestration for mlflow projects

`mlflow` `orchestration` `experiment-tracking` `DAG` `open-source`

📅 November 2023 - Present

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

`GitHub Actions` `Jinja2 Templates` `jsonschema` `open-source`

📅 February 2023 - Present

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,hmiladhia.github.io/cv.pdf.

🧊 splitter: Document Processing Tool

`Document Processing` `RAG` `open-source`

📅 December 2021 - Present

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

`Jupyter` `Databricks` `Zeppelin` `open-source`

📅 May 2021 - Present

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.

Data Wrangling

`pandas` `Polars` `PySpark` `🗄 SQL`

Cloud Technologies

`⊞ Azure ML` `Databricks`

LANGUAGES

Language	
Arabic	●●●●●
English	●●●●●
French	●●●●●
German	●○○○○

CERTIFICATES

🎓 **Machine Learning Engineering for Production (MLOps) Specialization**
DeepLearning.AI   📅 2022 2022

🎓 **Natural Language Processing Specialization**
DeepLearning.AI   📅 2024 2024

🎓 **Cloud-Native Development with Open-Shift 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