Dhia **Hmila**

📞 +33610201606 | ✉ dhia.hmila@dasquare.xyz | ⊠ hmiladhia.github.io |

🔘 hmiladhia | 🔗 dhia-hmila | Ⓜ Dhia Hmila | 📄 dhia-hmila

I am a Senior Data Scientist and I specialize in fraud/anomaly detection and document classification.

Professional experience

Data Science Team, AXA France   *Nanterre, France*
SENIOR DATA SCIENTIST   *2021 - Present*

- 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 PySpark 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.
- Automated the indexing of 70% of AXA France's incoming documents by developing a retrainable document classification AI System (exposed as an API).
- Reduced the workload of the help-desk by developing an LLM-based RAG system to answer insurance agents questions.
- Created a YAML-based templating system to streamline the creation of Azure ML pipelines.
- Built reusable Azure ML Pipelines (Continuous Training) to automatically retrain AI 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.

Data, Fraud, Waste & Abuse Team, AXA France   *Nanterre, France*
GRADUATE DATA SCIENTIST   *2020 - 2021*

- Engineered an advanced AI-driven Fraud, Waste, and Abuse Detection Tool that inspects different 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.

Shift Technology   *Paris, France*
DATA SCIENTIST   *2019 - 2020*

- 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.

U2IS - ENSTA Paris   *Palaiseau, France*
RESEARCH INTERNSHIP   *2018 - 2018*

- 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

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

mlflower: Lightweight orchestration tool for mlflow projects
HTTPS://GITHUB.COM/HMILADHIA/MLFLOWER   *2023 - Present*

`mlflow` `orchestration` `experiment-tracking`

Tool to extend mlflow projects with inbuilt multi step workflow orchestration

pysira: publish resumes in different formats
HTTPS://GITHUB.COM/HMILADHIA/PYSIRA *2023 - Present*

Github Actions   Jinja2 Templates   jsonschema

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, hmiladhia.github.io/fr/kendall or hmiladhia.github.io/resume.pdf

nbmanips: Split, merge and convert IPython Notebooks
HTTPS://PYPI.ORG/PROJECT/NBMANIPS/ *2021 - Present*

Jupyter   Databricks   Zeppelin

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
HTTPS://SPOT-LANGUAGE.ONRENDER.COM/ *2020 - 2020*

Classification   Flask   NLP   Lime

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.

Dmail
HTTPS://PYPI.ORG/PROJECT/DMAIL/ *2020 - 2020*

email   markdown   downloads>100k

A python package to write & send emails in markdown format
- Developed a python package to write and send emails in markdown format
- Used CI/CD piplines (Github Actions) to Lint, Test and publish the package to Pypi

DroNet
2019 - 2019

DroNet   CNN

DroNet for the forest environment
- Built a dataset from different sensors and a camera of a terrestrial mobile robot placed in a forest environment.
- Fine-tuned the DroNet neural network on the new dataset and tested the model in a Bebop drone

ENIT Robots competition
2017 - 2017

3 wheeled-robot   Arduino

Participation in ENIT Robots competition as team leader. Obtained 3rd place.

ENSIT Robots competition
2017 - 2017

OpenCV

Development in python of a trajectory recognition algorithm for a line following robot equipped with Raspberry Pi 3 and a camera.

Skills

Languages	Arabic | English | French | German
Machine Learning	mlflow | scikit-learn | Tensorflow | pytorch | trax | pyod | transformers | huggingface | pytorch-lightning
Computer Vision	OpenCV | torchvision | Vision Transformers | VGG16 | YOLO | Feature Matching | tesseract | easy-ocr | paddleOCR

Natural Language Processing	nltk | spaCy | Gensim | torchtext | langchain
Data Wrangling	pandas | pyspark | SQL | ETL
Programming Languages	Python | Rust | JavaScript | C/C++ | C#
MLOps	Github Actions | Azure Pipelines | Git | Linux | Docker | Kubernetes
Cloud	Azure ML | Databricks | Heroku
Web Scraping	selenium | requests | BeautifulSoup4
Web	HTML | CSS | JavaScript | fastapi | flask | Dash

Certificates

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