

Rami EL-WAZZI

Data Scientist — shipping ML models to production: time-series forecasting, NLP & LLM fine-tuning. Strong interest in MLOps & data engineering.

☎ +33 7 80 74 97 95 ✉ ramiwazzi@gmail.com 🌐 Rami-lwz 🌐 rami wazzi
🌐 French nationality 📍 Open to relocation: France & Europe

Skills

- **Languages:** Python, SQL, Scala
- **ML / Data Science:** time-series (SARIMA), clustering (DTW / DBA), NLP, LLM fine-tuning, XAI (LIME), scikit-learn
- **Big Data:** Spark, Hadoop, Hive, HDFS, Kafka, Spark Streaming
- **MLOps & Tools:** Streamlit, Dataiku, Talend, dbt, pytest, Git, design patterns
- **Cloud & Viz:** Azure (DP-100 certified), Qlik Sense

Professional Experience

Research Intern — Berner Fachhochschule (BFH)

Sep 2025 – Nov 2025

EU BIAS project — mitigating diversity biases of AI in the labour market

- Designed a **debiasing pipeline (GenWriter)** for **fairer CV screening in HR**, combining LLM-based sentence generalization with case-based reasoning (CBR).
- **Reduced gender bias by 35–37 %** (TPR gap) while retaining **97 % of classification performance**.
- Gender-labeling model (Norwegian sBERT + logistic regression) reaching **over 95 % accuracy**.
- Explainability analysis (**LIME / XAI**) confirming removal of gender cues from the model's decisions.
- **Co-author** of “Bias Mitigation in Hiring-Related NLP: Interactions Between Masking, Rewriting, and Adversarial Debiasing” (NLP4DH 2026).

Technologies: BERT / sBERT, LLMs (GPT, Mixtral, Llama, gpt-oss), NLP, XAI, Python

Data Scientist — French Ministry of Armed Forces

Oct 2022 – Sep 2025

Apprenticeship — Data & Digital Applications Centre (CDAN)

Data Engineering

- Built a **modular ETL pipeline in Python** (8 atomic stages, **pytest unit tests**) feeding a Big Data lake (Spark / Hadoop / Hive) and Qlik Sense dashboards — **the team's first industrialised Python pipeline, reused as a template** for later data flows.
- Designed an internal Big Data development environment (hybrid Datalake / Lakehouse) and handled data migration via Dataiku.

Data Science

- Forecasting model over **about 400 time-series** (hierarchical **DTW / DBA clustering**) to **anticipate operational readiness**, replacing a manual empirical method. **POC adopted as an official project**; 18 clusters for an MAE on par with individual-level prediction.
- Automated model training and monitoring through a **Streamlit interface** (metrics, *model drift*).

Technologies: Python, PySpark, Hadoop, Hive, SQL, Streamlit, Dataiku, Talend, Qlik Sense

Projects

- **falc'on (falc-on.fr):** data collection and **LLM fine-tuning**; still consulted on the model's training direction.
- **Imitation Learning:** AI agents solving Gymnasium environments, implemented **using only Python / numpy**.
- **Real-time IoT pipeline:** transfers hundreds of GB of IoT device data into a database within seconds (**Scala, Kafka, Spark Streaming**).

Education

- **Efrei Paris** — Engineering degree (MSc-level), Big Data & Machine Learning 2023 – 2025
Microsoft Certified: Azure Data Scientist Associate (DP-100)
- **SupDeVinci** — Master 1, Data Architecture 2022 – 2023
- **Université de Paris** — BSc in Computer Science 2019 – 2022

Languages & Interests

- **Languages:** French (native) • English (professional working proficiency — TOEIC 870) • Arabic (C2)
- **Interests:** Music (guitar for 12 years, composing) • Basketball • Game development (Unity)