



CONTACT

+33 7 80 74 97 95

ramiwazzi@gmail.com

Rami-lwz

rami wazzi

linktree

EDUCATION

- Engineering Degree in Big Data & Machine Learning

Efrei Paris, France, 2023 - 2025

- Certification DP100 Azure

Provided by Microsoft

Acquired March 2025, valid until March 2026

- Master 1 Data Architecture

SupDeVinci, France, 2022 – 2023

- Computer Science Bachelor

2019 – 2022, Université de Paris, France

- High School + Bac S spé maths

Lycée Abdel Kader, Lebanon, 2004 - 2019

Languages :

- French (fluent)
- Arabic (fluent)
- English (professional - TOEIC 870)

Hobbies:

- Music
- Basketball
- Game Development on Unity

PROFILE

I am an engineering graduate specializing in Big Data & Machine Learning with hands-on experience in data science, data engineering, and production ML systems, looking to specialize in Data Engineering. Background includes designing data pipelines, time series forecasting, and data architecture.

EXPERIENCE

Apprenticeship at the French MoD - Oct 2022 – Sept 2025

Keywords: PySpark, Talend, QlikSense, Dataiku, Time Series Forecasting, Hadoop, python, SQL

- Created and helped to design a data pipeline to aggregate data into a **data lake** from different sources to be used in a **QlikSense dashboard**. Pipeline was in python using pyspark and PgSQL scripts. Monitored the database using SQL and streamlit.
- **Automated the Training of time-series models** on a dataset that evolves monthly using streamlit to create an interface for monitoring models and displaying metrics on performance.
- Designed a development platform that best suited our team's needs in developing **Data Engineering / ML projects**.
- **I designed a proof of concept** that was presented to internal clients **that became a fully-fledged demanded project** that I designed with the help of my team.
- Around the end of the apprenticeship, I became a reference for the rest of my team for the feasibility of projects that my team could take on, **I was solicited for making such assessments**.
- Many other projects including one using **Dataiku** to migrate database and store backup tables as well as train models through the pipeline.

Many Academic Projects through my education - 2019 – 2024

Keywords: Azure, GCP, Reinforcement Learning, Computer Vision, NLP, LLM, open source LLM, Airflow, MLFlow

Those include 'end of course' projects, I was able to develop my skill across a broad range of technologies from web Developpement to fine-tuning LLM models etc.

- Scraping Website to create a dataset, then use Kibana to create a dashboard.
- **Creating "Imitation Learning" models using only python – numpy** to create AI agents able to solve games from gymnasium website.
- **Designing a whole data pipeline** that can (theoretically) pass hundreds of Gigabytes of data from IoT devices into a database in seconds using **Scala, Kafka and spark streaming**.
- Worked on the LLM model of <https://www.falc-on.fr> creating ways of gathering data for **fine-tuning LLMs** and then helping to fine-tune it. To this day I am still solicited for my opinion on the direction falc'on should take for its model training.
- I was also a **student-tutor giving Java OOP courses** to 2nd year students for 6 months.

Research Intern at Berner Fachhochschule - Sep 2025 – Dec 2025

Keywords: Bert, LLM, gpt-oss, mixtral, llama, NLP, eXplainable AI (XAI), python, numpy

- Currently working on the "BIAS european projet" which aims to reduce the bias in AI, including language models (LLMs and others). (<https://www.biasproject.eu/>) specifically, but usage not limited to, the labor market.
- My work there includes **explainable AI** methods used on those models to better understand how the social bias of human-made data (for example in HR) was included in "from the shelf" models. And methods to mitigate this bias.
- As of November 2025, I was able to **reduce the "bias" metric used by 60%** while not hurting accuracy of downstream tasks by applying state of the art methods proposed by myself after a careful study.

Internship at "Dar al handasah" - June – July 2021

Keywords: Hadoop, Spark, HDFS, Neo4J

- **Created a python library to facilitate the management of Data** in a Hadoop HDFS environment using pyspark and pyhive.
- Used Apache HUE to learn about how data is structured in the data lake.
- Created tools to gather statistics about partitions of data to find and delete duplicated data, sort files by a filter, etc.