

# Yann GUEGUEN

## PROFILE

Graduating Data Engineer in May, looking for a position focused on Cloud and AI.

## CONTACT

@ yann.gueguen.ad@gmail.com  
07 85 63 25 60  
ygueguen16u.github.io  
YGueguen16u  
Yann Gueguen

## SKILLS

### Languages and Scripting

- Python
- SQL
- R
- Linux, Bash, PowerShell
- React, HTML, CSS, JavaScript, Streamlit

### Big Data and Cloud Tools

- Apache Spark, Airflow, Hadoop
- AWS (S3, EC2, Lambda, ...)
- FastAPI,
- Databricks (basic knowledge)
- Dataiku
- PostgreSQL
- PowerBI

### AI Frameworks

- Machine Learning: Scikit-Learn, TensorFlow, PyTorch
- Deep Learning: Neural Networks, NLP, Computer Vision
- Hugging Face

### DevOps and Automation

- Docker
- CI/CD, Git, GitHub Actions
- Linux, Bash

## CERTIFICATIONS

- AWS Certified Solutions Architect – Associate (in progress)
- IBM Data Engineering Professional Certificate
- Neo4j Certified Professional
- Deep Learning Specialization – DeepLearning.AI

## LANGUAGES

- French (C2 – Native)
- English (C1 level)
- German (B1 level)

## HOBBIES

- Skiing (competition)
- Tennis
- Piano
- Street Work Out

## EXPERIENCE

### DATA ENGINEER (APPRENTICESHIP)

Thales Alenia Space | 01/2023 – 03/2025

- ◊ Designed and deployed ETL pipelines for test data from **SADM** (Solar Array Drive Mechanisms) and **SA** (Solar Arrays).
- ◊ Optimised data processing: table concatenation, moving average smoothing, and threshold filtering.
- ◊ Enhanced data quality using NLP, similarity algorithms, and Regex to detect and correct input errors.
- ◊ Built a FastAPI interface to automate file uploads to Dataiku and batch script execution (.bat).
- ◊ Automated CSV ingestion into Power BI using the Dataiku API for real-time dashboard updates.
- ◊ Reduced overall processing time (from several days to a few hours) through full pipeline automation.

### MAINTENANCE PROGRAMME ENGINEER

EDF | 09/2021 – 12/2022

- ◊ Oversaw Preventive Maintenance Programmes (PBMP) and technical inspections (ASG, SEC, LHP-LHQ, bolted assemblies) using historical data with SQL queries.
- ◊ Analysed inspection data and created trend reports in Power BI to improve maintenance planning.

## EDUCATION

### APPLIED MSC IN DATA SCIENCE

Data ScienceTech Institute | 02/2023 – 05/2025

- ◊ Data Science: statistics, machine learning, deep learning, time series, optimisation.
- ◊ Data Engineering: AWS, Big Data, SQL, MLOps, NoSQL, Python.
- ◊ **Solo Deep Learning Project:**
  - ◊ Created a custom food database via OpenFoodFacts scraping and cleaning.
  - ◊ Fine-tuned a Mixtral-8x7B model (Axolotl) for NLP-based intent recognition and structured log generation.
  - ◊ Backend: FastAPI, S3 storage, Airflow orchestration, CI/CD with GitHub Actions, deployed on Render.
  - ◊ Frontend: React-based food journal interface hosted on Vercel.
- ◊ Project : [Click here to view the project](#)

### ENGINEERING SCHOOL

ENSEM, Nancy | 09/2018 – 08/2021

- ◊ Applied mathematics, databases (SQL/NoSQL), graph theory, optimisation, dependability, signal processing.

### PREPARATORY CLASSES

Saint-Stanislas High School, Nantes | 09/2016 – 08/2018

- ◊ PCSI – PSI track (Maths, Physics, Engineering Sciences)

## TECHNICAL PROJECTS

### FOOD TRACKING APP (DE/ML)

- ◊ FastAPI backend simulating fake users and food intake logs (IoT-like time series).
- ◊ Data ingestion via API, AWS S3 storage, Python/DuckDB transformation, and Airflow orchestration.
- ◊ CI/CD pipelines using GitHub Actions.
- ◊ ML models (clustering, anomaly detection), visualised via dynamic Streamlit dashboards.
- ◊ Project : [Click here to view the project](#)