

# Yann GUEGUEN

## PROFIL

*Data Engineer* diplômé en mai, à la recherche d'un poste orienté Cloud et IA, avec un fort intérêt pour l'automatisation des traitements de données.

## CONTACT

@ yann.gueguen.ad@gmail.com

☎ 07 85 63 25 60

🌐 ygueguen16u.github.io

🐙 YGueguen16u

📄 Yann Gueguen

## COMPÉTENCES

### Langages et Scripting

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

### Outils Big Data et Cloud

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

### Frameworks IA

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

### DevOps et Automatisation

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

## CERTIFICATIONS

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

## LANGUES

- Français (C2 - Langue maternelle)
- Anglais (B2 – C1 en compréhension écrite)
- Allemand (B1 – intermédiaire)

## ACTIVITÉS EXTRA-PROFESSIONNELLES

- Ski (compétition)
- Tennis
- Piano
- Street Work out

## EXPERIENCE

### DATA ENGINEER EN ALTERNANCE

*Thales Alenia Space* | 01/2023 – 03/2025

- ◇ Développé des pipelines ETL pour les données issues des tests **SADM** (Solar Array Drive Mechanisms) et **SA** (Solar Arrays).
- ◇ Optimisé le traitement des données : concaténation de tables, réduction par moyenne mobile, et respect des seuils définis...
- ◇ Amélioré la qualité des données via NLP, algorithmes de similarité et Regex pour détecter et corriger les erreurs de saisie.
- ◇ Développé une interface Flask pour automatiser l'envoi des fichiers vers Dataiku et l'exécution des scripts batch (.bat).
- ◇ Automatisé l'intégration des fichiers CSV dans Power BI via l'API Dataiku pour une meilleure visualisation des données.
- ◇ Réduit le temps de traitement (de plusieurs jours à quelques heures) en automatisant l'ensemble du flux de données et en limitant les interventions manuelles.

### INGÉNIEUR D'ÉTUDES SUR LES PROGRAMMES DE MAINTENANCE

*EDF* | 09/2021 – 12/2022

- ◇ Supervisé les Programmes de Maintenance Préventive (PBMP) et les contrôles techniques (ASG, SEC, LHP-LHQ, assemblages boulonnés) en exploitant les données historiques via des requêtes SQL..
- ◇ Suivi des inspections et reporting des tendances via Power BI pour améliorer la planification des interventions.

## EDUCATION

### APPLIED MSC IN DATA SCIENCE

*Data ScienceTech Institute* | 02/2023 - 05/2025

- ◇ Data Science : statistiques, machine learning, deep learning, séries temporelles, optimisation.
- ◇ Data Engineering : AWS, Big Data, SQL, MLOps, NoSQL, Python.
- ◇ **Projet de Deep Learning en groupe (4 personnes) :**
  - ◇ Conception de modèles d'IA (NLP, Computer Vision) pour une application de suivi d'habitudes.
  - ◇ Responsable de la partie Data Engineering (Airflow, stockage cloud), NLP pour la reconnaissance d'intention (BERT/GPT) et la partie DevOps (CI/CD, Test, Cloud, validation des PR).

### ÉCOLE D'INGÉNIEUR

*ENSEM de Nancy* | 09/2018 – 08/2021

- ◇ Mathématiques appliquées, bases de données (SQL/NoSQL), optimisation, modélisation de systèmes, sûreté de fonctionnement, traitement du signal.

### CPGE SCIENTIFIQUE

*Lycée Saint-Stanislas de Nantes* | 09/2016 - 08/2018

- ◇ PCSI - PSI

## PROJETS TECHNIQUES

### FOOD TRACKING (DE/ML)

- ◇ API FastAPI simulant de faux utilisateurs et leurs activités alimentaires (données de type IoT).
- ◇ Collecte via API, stockage AWS S3, traitement Python/DuckDB, orchestration Airflow.
- ◇ CI/CD avec GitHub Actions.
- ◇ Modèles ML (clustering, détection d'anomalies), visualisation dynamique via Streamlit.