

Yann GUEGUEN

PROFIL

Data Engineer diplômé en mai, à la recherche d'un poste orienté Cloud et IA.

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
- React, HTML, CSS, JavaScript, Streamlit

Outils Big Data et Cloud

- Apache Spark, Airflow, Hadoop
- AWS (S3, EC2, Lambda, ...),
- FastAPI,
- 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 (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

- ◊ Conception de pipelines ETL pour exploiter les données de tests **SADM** (Solar Array Drive Mechanisms) et **SA** (Solar Arrays).
- ◊ Optimisation des traitements : concaténation de tables, moyennes mobiles, application de seuils techniques.
- ◊ Amélioration de la qualité des données via NLP, algorithmes de similarité et Regex pour détecter et corriger les erreurs de saisie.
- ◊ Développement d'une interface avec FastAPI pour automatiser l'envoi vers Dataiku et l'exécution des scripts batch (.bat).
- ◊ Intégration automatique des CSV dans Power BI via l'API Dataiku pour une visualisation fluide.
- ◊ Réduction drastique du temps de traitement (de plusieurs jours à quelques heures) grâce à l'automatisation complète du pipeline.

INGÉNIEUR D'ÉTUDES SUR LES PROGRAMMES DE MAINTENANCE

EDF | 09/2021 – 12/2022

- ◊ Supervision des Programmes de Maintenance Préventive (PBMP) et des contrôles techniques (ASG, SEC, LHP-LHQ, assemblages boulonnés), en exploitant les données historiques via requêtes SQL.
- ◊ Analyse des inspections et reporting des tendances sous Power BI pour optimiser 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 :**
 - ◊ Conception complète : scraping OpenFoodFacts, nettoyage, base alimentaire structurée.
 - ◊ Fine-tuning d'un Mixtral-8x7B (Axolotl) pour l'analyse NLP et la génération de logs.
 - ◊ Backend : FastAPI, S3, Airflow, CI/CD (GitHub Actions), déployé sur Render.
 - ◊ Frontend : app React type journal alimentaire, hébergée sur Vercel.
 - ◊ Code : [Cliquez ici pour voir le projet](#)

ÉCOLE D'INGÉNIEUR

ENSEM de Nancy | 09/2018 – 08/2021

- ◊ Mathématiques, bases de données (SQL/NoSQL), optimisation, théories des graphes, 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, nettoyage et traitement Python/DuckDB, orchestration Airflow.
- ◊ CI/CD avec GitHub Actions.
- ◊ Modèles ML (clustering, détection d'anomalies), visualisation dynamique via Streamlit.
- ◊ Project : [Cliquez ici pour voir le projet](#)