



Zaki Ilias

Data Engineer

Strasbourg, France (open to relocation)
+33 7 44 71 79 53
zakii.iliaas@gmail.com
iliaszaki.github.io
github.com/iliaszaki
in/ilias-zaki-b5841a226

Data Engineer completing a work-study contract, I design and automate end-to-end data platforms: 6 integrated API sources, 18 star-schema Delta tables and 5 pipelines orchestrated in Microsoft Fabric, now running in production. Graduating with an MSc in Big Data in 2026, I am seeking a **full-time Data Engineer position from September 2026**.

EXPERIENCE

SEP. 2025 – PRESENT

CTDI Europe (CTDI Group), Strasbourg, France
Data Engineer & Data Analyst (Work-Study)

- Optimized SOC data model: redesigned architecture, improved performance and reliability of cybersecurity data pipelines
- Architected Microsoft Fabric Lakehouse solution: Delta Lake storage, PySpark notebooks, orchestrated Data Factory pipelines
- Delivered production platform: 6 data sources, 18 Delta Lake tables, 5 pipelines end-to-end
- Built Power BI semantic models and reports for the SOC team: relationships, DAX measures, security KPIs, real-time alerts
- AI-driven reporting automation: risk classification (Power Automate, AI Builder), automatically generated alert investigation reports, MCP-based triage agent

MAY – AUG. 2025

CTDI GmbH (CTDI Group), Malsch, Germany
Data Engineer (Internship, 4 months)

- Built from scratch a production-ready multi-source SOC data pipeline, consolidating data from multiple cybersecurity APIs
- Delivered to production: 7 Delta Lake tables in Azure Fabric Lakehouse, fully automated archiving

PROJECTS

Agora Python, Apache Airflow, dbt, PostgreSQL

- End-to-end DE pipeline: synthetic e-commerce data (1,500–3,000 orders/day), Airflow orchestration, dbt transformations

Cloud Transcriber Azure Container Apps, Terraform, Docker, FastAPI, Whisper, GitHub Actions

- Audio transcription app on Azure: infrastructure as code, full CI/CD pipeline

ML Padel Project Python, scikit-learn, XGBoost, FFT, MFCC, Hilbert

- Classified ball impact zone, racket type and age from raw audio and vibration signals; 5 custom features; up to 100% test accuracy (KNN, Random Forest, SVM, XGBoost)

QuickQCM Python, PHP, OpenCV, EasyOCR, MySQL

- Web platform for MCQ creation, management and automated grading via image recognition; centralized result tracking

TECHNICAL SKILLS

DATA & BI Python, pandas, SQL, REST APIs, GraphQL, Delta Lake, dbt, Microsoft Fabric, Power BI, Power Automate, Azure Data Lake, Terraform
BIG DATA Apache Airflow, Spark (PySpark), Hadoop, Docker, Docker Compose, Azure Container Apps
ML & AI scikit-learn, XGBoost, CNN (MobileNetV2), Vision Transformer, Transfer Learning, R, Jupyter
DEVOPS FastAPI, Django, PHP, MySQL, GitHub Actions, CI/CD, OAuth2, OpenCV, EasyOCR

EDUCATION

2025 – 2026 **MSc Big Data**
IN PROGRESS
JUNIA ISEN, Lille
2024 – 2025 **Master 1, Big Data**
JUNIA ISEN, Lille
2023 – 2024 **1st Year Engineering Cycle**
TOP OF CLASS
ISEN, Rabat, Morocco
2021 – 2023 **Preparatory Cycle**
TOP OF CLASS
ISEN, Rabat, Morocco

CERTIFICATIONS

MICROSOFT **Power BI** : Harnessing Data, ETL, Data Modeling, Preparing Data for Analysis
STANFORD **Supervised Machine Learning** : Regression and Classification
IMPERIAL **Mathematics for ML** : Linear Algebra
MICHIGAN **Python** : Programming for Everybody, Data Structures

LANGUAGES

French native bilingual · **English** bilingual (TOEIC 850/990) · **Spanish** beginner