ALDORITH CHOUNA NGUEFA

Data Scientist — Machine Learning & Generative AI

Python · scikit-learn · statistics · NLP & RAG · clustering · model evaluation

Paris area (Cachan), France | +33 7 66 28 44 73 | chounanguefa005@gmail.com | portfolio.aldorith.com | LinkedIn | GitHub

French (native) · English C1 · German C1 · Available immediately · Permanent / freelance

PROFILE

Data scientist trained through an apprenticeship at **Société Générale** (Internal Audit): audit data analysis, clustering and a generative AI application in internal production. I care as much about **evaluation rigour** — data leakage, imbalance-aware metrics, business thresholds — as about the value delivered to teams.

PROFESSIONAL EXPERIENCE

Société Générale — Data & AI Scientist (apprenticeship), Audit & Inspection Générale Oct 2024 – Aug 2026

Automate the drafting of model-risk audit notes from large document sets.

- **Python** analysis and visualisation of audit datasets: challenged manual analyses, surfaced inconsistencies, built indicators for supervisory review.
- Hybrid **BM25 + FAISS** retrieval fused with **RRF**, section-level **GPT** generation, consolidation and **source attribution** for every passage, with human review kept in the loop.
- **HDBSCAN clustering** module (outlier handling) and business summaries with figure-level comparison across documents.
- **Outcome:** structured, sourced first draft in **minutes** for a process that took 2–3 months manually; iterated with managers; took part in strategic meetings with Data Science department leadership.

Le Comptoir d'Afrique (Afro Shop) — Freelance Full-Stack & AI Developer Sep 2026 – present

Digitising a grocery store: online shop, orders, pickup, delivery and store management.

- Gathered requirements with the owner and wrote the design file: user journeys, **SQL data model**, **OpenAPI contract**, target NestJS/PostgreSQL architecture and **RAG** assistant.
- **22 Vitest tests**, **7 Playwright E2E journeys**, axe-core accessibility checks; trained staff on the tool and cybersecurity basics. Version in acceptance testing; backend and payment to be integrated.

TECHNICAL PROJECTS

Phishing Detection — MLOps pipeline · code Oct 2026

- 11,055 rows, schema and drift validation (Kolmogorov-Smirnov), **duplicate-group splitting** (5,206 duplicates, zero train/test leakage), 5 models compared with cross-validation: **Random Forest test F1 0.959**.
- **FastAPI** and batch prediction, preprocessing + model serialised together, **Docker**, 8 pytest tests, **GitHub Actions** CI; **MLflow**/DagsHub tracking, initial **AWS ECR/EC2** deployment, S3 artefacts.

Bank churn prediction · code 2026

- Leak-free scikit-learn pipeline (preprocessing inside the model), stratified cross-validation, baseline, **PR-AUC** metric, threshold set on out-of-fold probabilities for a **0.75 target precision**; statistical tests of drivers.
- **Streamlit** app for single and CSV scoring, risk-ranked export, model card documenting limitations.

Prudential Evidence Lab — regulatory RAG assistant · demo Aug – Sep 2026

- **59 backend tests**, **30 golden evaluation cases** (100% citation precision, 0% false completeness on this set); hierarchical chunking benchmark: **MRR 0.603 → 0.693**.

SKILLS

Data science: Python, pandas, NumPy, scikit-learn, XGBoost, statistics, cross-validation, PR-AUC, threshold selection, Jupyter

NLP & generative AI: embeddings, RAG (BM25, FAISS, RRF), HDBSCAN/UMAP clustering, OpenAI/GPT, Gemini, LLM evaluation

Data & reporting: SQL, PostgreSQL, MongoDB, visualisation, Streamlit, FastAPI

Productionisation: MLflow, Docker, GitHub Actions, pytest, AWS (EC2, ECR, S3)

CERTIFICATIONS & TRAINING

Google Cloud Skills Boost — Machine Learning Operations (MLOps) Fundamentals · badge earned

Google Cloud Skills Boost — Develop Gen AI Apps with Gemini and Streamlit; Agent Governance and Security

Udemy — Deep Agent: Multi-Agent RAG with Gemini and LangChain (2026)

EDUCATION

ECE Paris — Engineering degree (MSc), Data & AI (2024–2026)

EFREI Paris — Computer science engineering track (2023–2024)

REFERENCES

Pierre-Michel Becquet — apprenticeship manager (2 years), Société Générale · pierre-michel.becquet@socgen.com

Jean Djeto — Senior Auditor, Audit & Inspection Générale, Société Générale · jean.djeto@socgen.com