ALDORITH CHOUNA NGUEFA

Data Scientist — Machine Learning & IA générative

Python · scikit-learn · statistiques · NLP & RAG · clustering · évaluation de modèles

Cachan (94), France | +33 7 66 28 44 73 | chounanguefa005@gmail.com | portfolio.aldorith.com | LinkedIn | GitHub

Français natif · Anglais C1 · Allemand C1 · Disponible immédiatement · CDI / freelance

PROFIL

Data scientist formé en alternance à **Société Générale** (Audit & Inspection Générale) : analyses de données d'audit, clustering et application d'IA générative en production interne. J'attache autant d'importance à la **rigueur de l'évaluation** — fuites de données, métriques adaptées au déséquilibre, seuils métier — qu'à la valeur livrée aux équipes.

EXPÉRIENCE PROFESSIONNELLE

Société Générale — Data & AI Scientist (alternance), Audit & Inspection Générale Oct. 2024 – août 2026

Automatiser la rédaction des notes de risque de modèle à partir de dossiers documentaires volumineux.

- Analyses et visualisations **Python** sur des jeux de données d'audit : challenge des analyses manuelles, détection d'incohérences, indicateurs pour la supervision.
- Recherche hybride **BM25 + FAISS** fusionnée par **RRF**, génération **GPT** par section, consolidation et **attribution de la source** de chaque passage ; validation humaine conservée.
- Module de **clustering HDBSCAN** (gestion des outliers) et de résumés métier avec comparaison chiffrée entre documents.
- **Résultat :** premier brouillon structuré et sourcé en **quelques minutes** pour un processus manuel de 2 à 3 mois ; itérations avec les managers ; participation aux réunions stratégiques du département Data Science.

Le Comptoir d'Afrique (Afro Shop) — Développeur Full-Stack & IA, freelance Sept. 2026 – présent

Digitaliser une épicerie : boutique en ligne, commandes, retrait, livraison et gestion du magasin.

- Recueilli les besoins avec la gérante et rédigé le dossier de conception : parcours, **modèle de données SQL**, **contrat OpenAPI**, architecture cible NestJS/PostgreSQL et assistant **RAG**.
- **22 tests** Vitest, **7 parcours E2E** Playwright, contrôle d'accessibilité axe-core ; formation de l'équipe à l'outil et aux bons réflexes de cybersécurité. Version en validation, backend et paiement à intégrer.

PROJETS TECHNIQUES

Phishing Detection — pipeline MLOps · code Oct. 2026

- 11 055 lignes, validation de schéma et de dérive (Kolmogorov-Smirnov), **découpage par groupes de doublons** (5 206 doublons, 0 fuite train/test), 5 modèles comparés en validation croisée : **Random Forest F1 test 0,959**.
- API **FastAPI** et prédiction par lots, prétraitement + modèle sérialisés ensemble, **Docker**, 8 tests pytest, CI **GitHub Actions** ; suivi **MLflow**/DagsHub, déploiement initial **AWS ECR/EC2**, artefacts S3.

Prédiction du churn bancaire · code 2026

- Pipeline scikit-learn sans fuite (prétraitement intégré), validation croisée stratifiée, baseline, métrique **PR-AUC**, seuil choisi sur probabilités out-of-fold pour une **précision cible de 0,75** ; tests statistiques des facteurs.
- Application **Streamlit** de scoring individuel et CSV, export des clients triés par risque, fiche modèle documentant les limites.

Prudential Evidence Lab — assistant RAG réglementaire · démo Août – sept. 2026

- **59 tests** backend, **30 cas d'évaluation** dorés (précision des citations 100 %, fausse complétude 0 % sur ce jeu) ; benchmark de chunking hiérarchique : **MRR 0,603 → 0,693**.

COMPÉTENCES

Data science : Python, pandas, NumPy, scikit-learn, XGBoost, statistiques, validation croisée, PR-AUC, choix de seuil, Jupyter

NLP & IA générative : embeddings, RAG (BM25, FAISS, RRF), clustering HDBSCAN/UMAP, OpenAI/GPT, Gemini, évaluation LLM

Données & restitution : SQL, PostgreSQL, MongoDB, visualisation, Streamlit, FastAPI

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

CERTIFICATIONS & FORMATIONS

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

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

Udemy — Deep Agent : Multi-Agent RAG avec Gemini et LangChain (2026)

FORMATION

ECE Paris — Diplôme d'ingénieur Data & IA (2024–2026)

EFREI Paris — Cycle ingénieur informatique (2023–2024)

RÉFÉRENCES

Pierre-Michel Becquet — manager d'alternance (2 ans), Société Générale · pierre-michel.becquet@socgen.com

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