ALDORITH CHOUNA NGUEFA

Ingénieur Machine Learning

Python · scikit-learn · embeddings & recherche vectorielle · modèles multimodaux · RAG

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

Ingénieur ML (2 ans à **Société Générale**) : de la donnée au modèle servi. Recherche hybride et clustering en production interne, pipelines de classification sans fuite, **recherche multimodale** par embeddings et benchmarks de récupération. Je mesure avant d'affirmer.

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.

- 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.
- Analyses et visualisations **Python** sur des jeux de données d'audit : challenge des analyses manuelles, détection d'incohérences, indicateurs pour la supervision.
- **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**.
- Développé le frontend **Next.js 16 / React 19 / TypeScript** (catalogue, panier, commande, compte, assistant de démonstration) sur des ports de services typés depuis l'OpenAPI.

PROJETS TECHNIQUES

MultiPix — recherche photo multimodale (projet de fin d'études, binôme) · code 2025 – 2026

- Mon périmètre : recherche texte → image et image → image, RAG multimodal hybride, clustering et orchestration des pipelines d'indexation : API FastAPI, PostgreSQL, file **Redis**, worker d'embeddings **TinyCLIP** (poids figés SHA-256), **Qdrant** ; fusion plein texte + vecteurs par **RRF**, reranking optionnel par VLM local.
- **34 tests** pytest, smoke test de bout en bout, CI GitHub Actions, Docker Compose ; benchmarks OpenCLIP / BLIP / HDBSCAN sur 5 000 images COCO.

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.

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

- RAG Solvabilité II : Docling/PyMuPDF, **BM25 + embeddings Gemini + RRF**, porte de preuves déterministe (entité, période, champs requis), statut COMPLETE / PARTIAL / CONFLICT / NOT_FOUND et une citation par affirmation.
- **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

Machine learning : scikit-learn, XGBoost, validation croisée, contrôle des fuites, PR-AUC, HDBSCAN/UMAP

Embeddings & retrieval : TinyCLIP, OpenCLIP, BLIP, Gemini embeddings, FAISS, Qdrant, BM25, RRF, MRR, reranking

Serving & données : Python, FastAPI, PostgreSQL, Redis, workers asynchrones, SQL

Livraison : Docker, GitHub Actions, pytest, MLflow, Prometheus, Grafana, 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