ALDORITH CHOUNA NGUEFA

Ingénieur IA — RAG & systèmes agentiques

Python · LLM · RAG hybride · LangGraph · MCP · évaluation & sécurité des LLM

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 IA, deux ans d'alternance à **Société Générale** où j'ai livré une application RAG utilisée par l'Audit pour rédiger des notes de risque de modèle. Je conçois des **applications LLM traçables** : recherche hybride, agents, contrôle des citations, évaluation et défenses contre l'injection de prompt, du cadrage métier à la mise en production.

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.

- Conçu et développé une **application d'IA générative** de bout en bout : ingestion de rapports PDF, Word et Excel, extraction et préparation documentaire, backend **Python/FastAPI** et interface web.
- 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**.
- 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

Folio — recherche financière multi-agent · démo 2026

- **Chat RAG contrôlé en 8 étapes** sur 28 rapports SEC : BM25, génération Gemini en JSON strict, vérification des citations, montants, années et entreprises ; abstention plutôt qu'invention.
- **4 agents LangGraph** (planificateur, chercheur, vérificateur, rédacteur) avec transferts signés HMAC ; données de marché via **MCP** Yahoo Finance derrière une passerelle en lecture seule.
- React 19, Node.js 22, worker Python 3.12, Docker durci ; observabilité **Prometheus, Grafana, LangSmith** ; Oracle Cloud + Cloudflare. **250+ tests** ; red team : **165 attaques, 0 réussie**, 0 faux positif sur 170 cas légitimes.

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**.

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.

COMPÉTENCES

IA générative : RAG hybride (BM25, FAISS, Qdrant, RRF), LangGraph, LangChain, MCP, OpenAI/GPT, Gemini, embeddings, Docling

Évaluation & observabilité : jeux de cas dorés, précision des citations, red teaming, LangSmith, Prometheus, Grafana

Sécurité applicative : OWASP Top 10 LLM, défense contre l'injection de prompt, HMAC entre agents, CSRF, 2FA, CSP, conteneurs durcis

Backend & données : Python, FastAPI, Pydantic, Node.js, SQL, PostgreSQL, SQLite, REST/OpenAPI

Livraison : Docker, GitHub Actions, pytest, AWS (EC2, ECR, S3), Oracle Cloud, Cloudflare, Git

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