ALDORITH CHOUNA NGUEFA

Machine Learning Engineer

Python · scikit-learn · embeddings & vector search · multimodal models · RAG

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

ML engineer (2 years at **Société Générale**): from data to served model. Hybrid retrieval and clustering in internal production, leakage-free classification pipelines, **multimodal search** with embeddings and retrieval benchmarks. I measure before I claim.

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.

- 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.
- **Python** analysis and visualisation of audit datasets: challenged manual analyses, surfaced inconsistencies, built indicators for supervisory review.
- **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.
- Built the **Next.js 16 / React 19 / TypeScript** frontend (catalogue, cart, checkout, account, demo assistant) on service ports typed from the OpenAPI spec.

TECHNICAL PROJECTS

MultiPix — multimodal photo search (final-year project, team of two) · code 2025 – 2026

- My scope: text → image and image → image search, hybrid multimodal RAG, clustering and indexing pipeline orchestration: FastAPI API, PostgreSQL, **Redis** queue, **TinyCLIP** embedding worker (SHA-256-pinned weights), **Qdrant**; full-text + vector **RRF** fusion, optional local VLM reranking.
- **34 pytest tests**, end-to-end smoke test, GitHub Actions CI, Docker Compose; OpenCLIP / BLIP / HDBSCAN benchmarks on 5,000 COCO images.

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.

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

- Solvency II RAG: Docling/PyMuPDF, **BM25 + Gemini embeddings + RRF**, deterministic evidence gate (entity, period, required fields), COMPLETE / PARTIAL / CONFLICT / NOT_FOUND status and one citation per claim.
- **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

Machine learning: scikit-learn, XGBoost, cross-validation, leakage control, PR-AUC, HDBSCAN/UMAP

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

Serving & data: Python, FastAPI, PostgreSQL, Redis, async workers, SQL

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