ALDORITH CHOUNA NGUEFA

Machine-Learning-Ingenieur

Python · scikit-learn · Embeddings & Vektorsuche · multimodale Modelle · RAG

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

Französisch (Muttersprache) · Englisch C1 · Deutsch C1 · Sofort verfügbar · Festanstellung / Freelance

PROFIL

ML-Ingenieur (2 Jahre bei der **Société Générale**): von den Daten bis zum bereitgestellten Modell. Hybride Suche und Clustering im internen Betrieb, leakage-freie Klassifikations-Pipelines, **multimodale Suche** mit Embeddings und Retrieval-Benchmarks. Ich messe, bevor ich behaupte.

BERUFSERFAHRUNG

Société Générale — Data & AI Scientist (duales Studium), Konzernrevision Okt. 2024 – Aug. 2026

Automatisierung der Modellrisiko-Berichte der Revision auf Basis umfangreicher Dokumentenbestände.

- Hybride Suche **BM25 + FAISS** mit **RRF**-Fusion, abschnittsweise **GPT**-Generierung, Konsolidierung und **Quellenangabe** für jede Passage; menschliche Prüfung beibehalten.
- Modul für **HDBSCAN-Clustering** (Ausreißerbehandlung) und fachliche Zusammenfassungen mit Zahlenvergleich zwischen Dokumenten.
- **Python**-Analysen und Visualisierungen von Revisionsdaten: manuelle Analysen hinterfragt, Inkonsistenzen aufgedeckt, Kennzahlen für die Aufsicht.
- **Ergebnis:** strukturierter, belegter erster Entwurf in **wenigen Minuten** statt 2–3 Monaten manueller Arbeit; Iterationen mit den Führungskräften, Teilnahme an strategischen Sitzungen der Data-Science-Abteilung.

Le Comptoir d'Afrique (Afro Shop) — Full-Stack- & KI-Entwickler, Freelance Sept. 2026 – heute

Digitalisierung eines Lebensmittelgeschäfts: Onlineshop, Bestellungen, Abholung, Lieferung, Filialverwaltung.

- Anforderungen mit der Inhaberin erhoben, Konzeptdossier erstellt: Nutzerwege, **SQL-Datenmodell**, **OpenAPI-Vertrag**, Zielarchitektur NestJS/PostgreSQL und **RAG**-Assistent.
- Frontend mit **Next.js 16 / React 19 / TypeScript** (Katalog, Warenkorb, Bestellung, Konto, Demo-Assistent) auf Service-Ports mit aus OpenAPI generierten Typen.

TECHNISCHE PROJEKTE

MultiPix — multimodale Fotosuche (Abschlussprojekt, Zweierteam) · Code 2025 – 2026

- Mein Bereich: Suche Text → Bild und Bild → Bild, hybrides multimodales RAG, Clustering und Orchestrierung der Indexierung: FastAPI-API, PostgreSQL, **Redis**-Warteschlange, **TinyCLIP**-Embedding-Worker (per SHA-256 fixierte Gewichte), **Qdrant**; **RRF**-Fusion aus Volltext und Vektoren, optionales Reranking per lokalem VLM.
- **34 pytest-Tests**, End-to-End-Smoke-Test, GitHub-Actions-CI, Docker Compose; Benchmarks OpenCLIP / BLIP / HDBSCAN auf 5.000 COCO-Bildern.

Phishing Detection — MLOps-Pipeline · Code Okt. 2026

- 11.055 Zeilen, Schemaprüfung, Drift-Test (Kolmogorov-Smirnov), **Split nach Duplikatgruppen** (5.206 Duplikate, kein Leakage), 5 Modelle mit Kreuzvalidierung: **Random Forest, Test-F1 0,959**.
- **FastAPI** und Batch-Vorhersage, KNN-Imputation und Modell gemeinsam serialisiert, **Docker**, 8 pytest-Tests, **GitHub-Actions**-CI; **MLflow**/DagsHub, erstes Deployment auf **AWS ECR/EC2**, Artefakte in S3.

Prudential Evidence Lab — regulatorischer RAG-Assistent · Demo Aug. – Sept. 2026

- Solvency-II-RAG: Docling/PyMuPDF, **BM25 + Gemini-Embeddings + RRF**, deterministische Nachweisprüfung (Einheit, Zeitraum, Pflichtfelder), Status COMPLETE / PARTIAL / CONFLICT / NOT_FOUND, ein Zitat je Aussage.
- **59 Backend-Tests**, **30 Goldfälle** (100 % Zitatgenauigkeit, 0 % falsche Vollständigkeit auf diesem Datensatz); Benchmark hierarchisches Chunking: **MRR 0,603 → 0,693**.

KOMPETENZEN

Machine Learning: scikit-learn, XGBoost, Kreuzvalidierung, Leakage-Kontrolle, PR-AUC, HDBSCAN/UMAP

Embeddings & Retrieval: TinyCLIP, OpenCLIP, BLIP, Gemini-Embeddings, FAISS, Qdrant, BM25, RRF, MRR, Reranking

Serving & Daten: Python, FastAPI, PostgreSQL, Redis, asynchrone Worker, SQL

Auslieferung: Docker, GitHub Actions, pytest, MLflow, Prometheus, Grafana, AWS (EC2, ECR, S3)

ZERTIFIKATE & WEITERBILDUNG

Google Cloud Skills Boost — Machine Learning Operations (MLOps) Fundamentals · Badge erworben

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

Udemy — Deep Agent: Multi-Agent RAG mit Gemini und LangChain (2026)

AUSBILDUNG

ECE Paris — Ingenieurdiplom (Master) Data & KI (2024–2026)

EFREI Paris — Ingenieurstudium Informatik (2023–2024)

REFERENZEN

Pierre-Michel Becquet — Betreuer im dualen Studium (2 Jahre), Société Générale · pierre-michel.becquet@socgen.com

Jean Djeto — Senior Auditor, Konzernrevision, Société Générale · jean.djeto@socgen.com