ALDORITH CHOUNA NGUEFA

KI-Ingenieur — RAG & agentische Systeme

Python · LLMs · hybrides RAG · LangGraph · MCP · Evaluation & Sicherheit von LLMs

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

KI-Ingenieur mit zwei Jahren dualem Studium bei der **Société Générale**, wo ich eine RAG-Anwendung für die Modellrisiko-Berichte der Revision geliefert habe. Ich baue **nachvollziehbare LLM-Anwendungen**: hybride Suche, Agenten, Zitatprüfung, Evaluation und Schutz vor Prompt-Injection, von der fachlichen Abstimmung bis zum Betrieb.

BERUFSERFAHRUNG

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

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

- Konzeption und Entwicklung einer **generativen KI-Anwendung** von Ende zu Ende: Aufnahme von PDF-, Word- und Excel-Berichten, Extraktion und Aufbereitung, **Python/FastAPI**-Backend und Weboberfläche.
- 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.
- **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

Folio — Multi-Agent-Finanzrecherche · Demo 2026

- **RAG-Chat in 8 festen Schritten** über 28 SEC-Berichte: BM25, Gemini mit striktem JSON, Prüfung von Zitaten, Beträgen, Jahren und Unternehmen; Enthaltung statt Erfindung.
- **4 LangGraph-Agenten** (Planer, Rechercheur, Prüfer, Autor) mit HMAC-signierten Übergaben; Marktdaten über **MCP** Yahoo Finance hinter einem nur lesenden Gateway.
- React 19, Node.js 22, Python-3.12-Worker, gehärtetes Docker; Observability mit **Prometheus, Grafana, LangSmith**; Oracle Cloud + Cloudflare. **250+ Tests**; Red Team: **165 Angriffe, 0 erfolgreich**, 0 Fehlalarme bei 170 legitimen Fällen.

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**.

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.

KOMPETENZEN

Generative KI: hybrides RAG (BM25, FAISS, Qdrant, RRF), LangGraph, LangChain, MCP, OpenAI/GPT, Gemini, Embeddings, Docling

Evaluation & Observability: Goldfälle, Zitatgenauigkeit, Red Teaming, LangSmith, Prometheus, Grafana

Anwendungssicherheit: OWASP Top 10 für LLMs, Abwehr von Prompt-Injection, HMAC zwischen Agenten, CSRF, 2FA, CSP, gehärtete Container

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

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

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