ALDORITH CHOUNA NGUEFA

Data Scientist — Machine Learning & generative KI

Python · scikit-learn · Statistik · NLP & RAG · Clustering · Modellevaluation

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

Data Scientist, ausgebildet im dualen Studium bei der **Société Générale** (Konzernrevision): Analyse von Revisionsdaten, Clustering und eine generative KI-Anwendung im internen Betrieb. Mir ist die **Strenge der Evaluation** – Leakage, unbalancierte Metriken, fachliche Schwellen – ebenso wichtig wie der Nutzen für die Teams.

BERUFSERFAHRUNG

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

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

- **Python**-Analysen und Visualisierungen von Revisionsdaten: manuelle Analysen hinterfragt, Inkonsistenzen aufgedeckt, Kennzahlen für die Aufsicht.
- 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.
- **22 Vitest-Tests**, **7 Playwright-E2E-Abläufe**, axe-core-Barrierefreiheit; Schulung des Teams zu Werkzeug und Cybersicherheit. In der Abnahme; Backend und Zahlung folgen.

TECHNISCHE PROJEKTE

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.

Vorhersage der Bankkunden-Abwanderung · Code 2026

- Leakage-freie scikit-learn-Pipeline (Vorverarbeitung im Modell), stratifizierte Kreuzvalidierung, Baseline, Metrik **PR-AUC**, Schwelle auf Out-of-Fold-Wahrscheinlichkeiten für eine **Zielpräzision von 0,75**; statistische Tests der Einflussfaktoren.
- **Streamlit**-App für Einzel- und CSV-Scoring, nach Risiko sortierter Export, Model Card mit Grenzen.

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

- **59 Backend-Tests**, **30 Goldfälle** (100 % Zitatgenauigkeit, 0 % falsche Vollständigkeit auf diesem Datensatz); Benchmark hierarchisches Chunking: **MRR 0,603 → 0,693**.

KOMPETENZEN

Data Science: Python, pandas, NumPy, scikit-learn, XGBoost, Statistik, Kreuzvalidierung, PR-AUC, Schwellenwahl, Jupyter

NLP & generative KI: Embeddings, RAG (BM25, FAISS, RRF), HDBSCAN/UMAP-Clustering, OpenAI/GPT, Gemini, LLM-Evaluation

Daten & Reporting: SQL, PostgreSQL, MongoDB, Visualisierung, Streamlit, FastAPI

Industrialisierung: MLflow, Docker, GitHub Actions, pytest, 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