Alain Lorenz

CTO & Mitgründer — KI & MLOps

Zwischen Genf und Lyon · al@inlore.nz · ala.inlore.nz/de · linkedin.com/in/alain-lorenz

Sie sind ein Mensch? Hier geht es zum visuellen Lebenslauf und zur Website: Visueller Lebenslauf · ala.inlore.nz

PROFIL

Als Mitgründer und CTO habe ich Teams von bis zu zehn Personen geleitet, ohne die Tastatur loszulassen: Ich bin immer noch der größte Contributor der API, die ich 2016 gestartet habe. Seit 2023 ist KI in Produktion mein Hauptterrain — Modelle trainieren, deployen, bereitstellen und optimieren, die echten Nutzern antworten. Meine Lieblingsthemen: LLMs, RAG und Fine-Tuning. Und der Papierkram, der die Firma am Laufen hält — DSGVO, DAC7, Steuern, Innovationssteuergutschrift — läuft auch über mich. Alles zwischen Genf und Lyon.

BERUFSERFAHRUNG

Technischer Mitgründer · CTO — JobyPepper

seit 2016 · jobypepper.com

Die Plattform, die Unternehmen und Freelancer für Einsätze zusammenbringt.

- Web, Mobile, API und Infrastruktur: das gesamte Produkt-Ökosystem.
- Zwei Mobile Apps in den iOS- und Android-Stores veröffentlicht, vom Build bis zum Store-Eintrag: Versionen, Einreichung, Compliance-Anpassungen, Texte.
- Zahlungs-Dispatching statt bloß eines „Bezahlen"-Buttons: Der Kunde zahlt einmal, die Plattform erstellt die Rechnung jedes Freelancers, verteilt die Mittel und löst jede Auszahlung aus. Automatisch.
- KI-Agent für Kandidaten-Sourcing und -Matching, angebunden über MCP.

Technischer Mitgründer · CTO — Ozly

2023 · pausiert · hello.ozly.ai

KI für die Jobsuche.

- Komplette Pipeline: Ingestion aus mehreren Quellen, CamemBERT-Klassifizierung, Vektorindex.
- Semantische Suche und konversationelles RAG, in einem einbettbaren Widget.
- Optimierte GPU-Inferenz: Kosten durch vier geteilt.

MLOPS

- **Sammeln:** Geplante Pipelines, die Daten aus mehreren Quellen zusammenführen, bereinigen und deduplizieren.
- **Trainieren:** Aufbau von Datensätzen, Fine-Tuning von Klassifikatoren nach Branche und Vertragsart.
- **Verpacken:** Versionierte Modelle, CUDA-Docker-Images, veröffentlicht in einer Container-Registry.
- **Bereitstellen:** GPU-Inferenz, Embedding-Dienst auf Kubernetes, Vektorsuche.
- **Betreiben:** CI/CD, Observability auf allen Diensten, laufende Optimierung: Embeddings auf 512 Dimensionen reduziert, ein Speicherleck aufgespürt und behoben.

FÜHREN

- Technische Leitung eines Teams von bis zu zehn Personen.

- Architekturentscheidungen, Lieferplanung, Abwägungen zwischen Produkt und Technik.
- Erstellung von Anträgen zur Innovationsförderung.
- Aus einem operativen Bedarf ein Werkzeug machen, aus einem KI-Experiment einen laufenden Dienst.

DAS UNTERNEHMEN AM LAUFEN HALTEN

- DSGVO-Konformität und die dazugehörigen Meldungen.
- DAC7-Exporte und -Meldungen, die Steuermeldepflicht für Plattformen der Sharing Economy.
- Unternehmenssteuern und Meldepflichten.
- Begleitung der Projekte, die für die französische Innovationssteuergutschrift (CII) in Frage kommen.

NEBENBEI

Kundenprojekt — p.oint

2026

Anwesenheitserfassung für Schulen und Jugendeinrichtungen: Namen auf dem Tablet antippen, „Ich bin da“, „Ich gehe“.

- Tablet-Terminal als PWA, das auch ohne Netz weiter erfasst und synchronisiert, sobald die Verbindung zurück ist.
- Verwaltung für Betreuer: Sessions, Statistiken, Excel-kompatibler CSV-Import/-Export, Tablets per QR-Code gekoppelt.
- Schlüsselfertig ausgeliefert: Node- + SQLite-API, Debian-Installation, HTTPS und tägliche Backups.

Persönliches Projekt — Origin AI

2026

Videos echter Arbeitsplätze, verwandelt in Trainingsdaten für die Robotik.

- Fragmentierter Video-Upload, ffmpeg-Verarbeitung, Tagging per LLM.
- Automatisches Verpixeln von Gesichtern auf GPU, Pipeline zur Datensatzerstellung.

Side Project — EV Comparator

2026

Ab wie vielen Kilometern wird das Elektroauto günstiger?

- Tägliche Open-Data-Ingestion, Kostenmodell pro Kilometer.
- Wiederverkaufswert als Schätzung ausgewiesen, nie als Wahrheit.

KOMPETENZEN

KI & MLOps: LLMs, RAG, Embeddings und Vektorsuche, Fine-Tuning (CamemBERT), Triton Inference Server, CUDA, MCP-Server, LangChain

Deployment: Docker (CUDA-Images, Compose mit GPU), Kubernetes, CI/CD (CircleCI), Nginx, Sentry

Backend: Node.js / Express, Python, RabbitMQ, MongoDB, Redis, WebSocket

Frontend & Mobile: React, React Native / Expo, Veröffentlichung für iOS und Android

Management: Technische Leitung (bis zu 10 Personen), DSGVO, DAC7, Steuern, CII

SPRACHEN

Französisch, Deutsch, Englisch