

AI Engineer & Delivery Lead

Full-stack engineer for ten years, building AI systems since 2023: I take proofs of concept to production, keep them running, and bring the teams that inherit them up to speed. **Cross-border commuter, five years working in Luxembourg** at Redspher.

Zoufftgen, France – on the Luxembourg border · cross-border commuter

3,300+ commits Contributions over the last 12 months across 33 repositories – still hands-on daily	~60% Share of the AI SmartTalk product codebase written by me – from first PoC to production	11 modules Generative-AI programme designed and delivered to IT teams – developer to executive level	5 teams Cross-functional teams led at Redspher, 2 Product Owners managed, 4-product backlog
--	--	--	---

PROFILE

I write the code and I run the delivery. Hands-on: RAG pipelines on PostgreSQL and pgvector, agentic workflows, MCP servers, self-hosted open-source models – containerised and shipped through CI/CD with unit and E2E coverage. Not only LLMs: I have also shipped classical ML in production, price-prediction models and a bin-packing optimiser running inside a purchasing pipeline. On the delivery side: five years as Product Manager in Luxembourg, a four-product backlog arbitrated with the CPO and CEO, two Product Owners managed and five cross-functional teams led. **What I do best is turn a proof of concept into something an organisation can actually operate** – documented architecture, tested pipelines, reproducible deployments, and engineers trained to run it without me.

SKILLS

AI & ML engineering

Python · RAG (pgvector, embeddings, chunking, re-ranking) · LangChain & LangGraph · MCP servers and clients · agentic workflows · Anthropic & OpenAI APIs · prompt engineering & evals · self-hosted models (Ollama, vLLM, llama.cpp) · machine learning with TensorFlow

Data pipelines & management

Data modelling and preparation for data-warehouse feeds (BigQuery), alongside data analysts · ingestion connectors (SharePoint, Google Drive, Jira, Airtable, CMS) · extraction, chunking, embedding and incremental re-indexing · PostgreSQL & pgvector · Redis · RabbitMQ · schema and contract validation (Zod, JSON Schema) · PII anonymisation before inference · EU hosting and GDPR constraints

Engineering & infrastructure

TypeScript · React · Next.js · Node.js · Python · PHP (Symfony, Laravel) · C# · REST and typed API contracts · Docker · Kubernetes · GitHub Actions CI/CD · unit and E2E test suites (Playwright, Cypress) · nginx with automatic TLS · monitoring · on-premise and self-hosted deployments

Delivery & enablement

Use-case backlog ownership and prioritisation · taking PoCs and MVPs to production · solution architecture and procedure documentation · Scrum & Kanban · Jira · workshop facilitation and hands-on training · mentoring engineering teams · reporting to CPO and C-level · managing Product Owners

EXPERIENCE

- **AI Trainer – generative AI for engineering teams** Jan 2026 – Present

Wavia · France – on site & remote

 - Design and deliver hands-on generative-AI training to engineering organisations – developers, QA and product – in formats from 2-hour framing sessions to full days of lab work, on site or remote.
 - 11-module curriculum built from scratch:** 400+ slides, live demos, exercises and graded assessments. The decks are **generated from code and versioned in Git**, so every session is reproducible and stays current as the tooling moves.
 - Modules cover AI-assisted coding (Claude Code, Cursor, Windsurf, Atlassian Rovo), **MCP servers and clients**, AI-driven test automation for web and mobile, **self-hosted open-source models**, and writing functional specifications with AI.
 - Each module ends on a graded assessment, so adoption is measured rather than assumed. Clients include **Pixine** (a full development team, 2 × 7h) and **Proximit**.
- **Co-Founder & CEO – AI platform** Feb 2024 – Present

AI SmartTalk · France – remote

 - Wrote the initial proof of concept myself**, then brought in a co-founder and industrialised it into a production platform. I have written **roughly 60% of the product code** and still ship daily – 3,300+ contributions across 33 repositories over the last 12 months.
 - RAG-as-a-Service architecture:** Next.js and Node.js services with Python workers, PostgreSQL + pgvector, Redis, LangChain / LangGraph, Anthropic and OpenAI APIs – containerised with Docker, deployed on Kubernetes through GitHub Actions CI/CD, with unit and E2E suites gating every release.

- **Ingestion pipelines** feeding the vector store from the systems clients already run – SharePoint, Google Drive, Jira, Airtable, CMS platforms: extraction, chunking, embedding, re-ranking and incremental re-indexing, so assistants answer from existing enterprise knowledge instead of a copy of it. I also **structured the platform's own data for the data warehouse** – conversation, usage and knowledge-base schemas prepared for analytics.
- **Operating it in production:** answer-quality evaluation, retrieval and prompt tuning, latency and token-cost optimisation, monitoring and incident handling on live assistants serving **37 languages**. Integration surface: REST API, public npm SDK and **bidirectional MCP** (tool server plus OAuth client of third-party servers).
- **Deployment under constraint:** EU hosting, GDPR alignment, on-premise and fully self-hosted installations, and **self-hosted open-source models** (Ollama, vLLM) where data cannot leave the client's infrastructure.

Product Manager – 4-product portfolio

May 2018 – Jun 2023

Redspher · Luxembourg

Five years in Luxembourg, promoted from consultant to Product Manager in a logistics-tech scale-up. **Reported to the CPO and the CEO, managed 2 Product Owners** and **led 5 cross-functional teams** across a **portfolio of 4 products**. **Stayed hands-on throughout:** full-stack development (PHP / Symfony, Node.js, React), REST API design, architecture decisions and **machine learning in Python and TensorFlow**. Joined via Adneom (2018–2019), then hired directly.

- **Automations** (2022–2023) – automated **40% of road-vehicle purchasing**: **machine-learning models in Python and TensorFlow** predicting the average purchase price for a given flow and required vehicle, combined with a **bin-packing** solver for load consolidation, wired straight into the ordering pipeline. **5 cross-functional teams and 2 Product Owners** under my responsibility.
- **Shipper Portal** (2021–2023) – B2B ordering platform built from zero, now used daily by **1,000+ users**; **60% of the customer base** migrated off phone ordering. Introduced **automated E2E testing** into the delivery pipeline – **hotfixes down 30%**. Trained a QA analyst who became the company's lead QA.
- **Oscar** (2021–2022) – shipment monitoring platform adopted by **500+ internal users**; structured user interviews cut a large slice of planned development before it was written. Onboarded a Product Owner to run it independently.
- **Skoreway** (2020–2023) – scaled and stabilised a B2B SaaS: production incidents from **10 tickets a day to 10 a month**.
- **Data & analytics** – worked hand in hand with the data analysts: modelled and prepared the application data feeding the group's **BigQuery data warehouse** across the products I owned – schemas, extraction and data quality.

AI Consultant – CAT tool

Jun – Sep 2024

Acolad · Paris – remote

- Integrated LLM capabilities into a Computer-Assisted Translation tool for a large language-services group – production LLM work inside an established enterprise codebase, with contributions pushed back to its open-source components.

Earlier

2016 – 2024

Freelance consulting · Webforger web agency · Juxta

Freelance engineering for **TextMaster** (Symfony 5 middleware on the Akeneo API – PHP 8, Docker, CI/CD), **Viaweb** (Docker / Kubernetes containerisation, new API services) and an NDA client (AI search on pgvector + OpenAI embeddings, E2E test automation on Angular 7 – the project that started my move into AI). Founded the **Webforger** web agency, still running as a side business. Junior analyst-developer at **Juxta**, which had no deployment tooling at all: I built its **CI/CD pipeline and multi-instance release-management tool** (C#, JavaScript).

OPEN SOURCE

AnonDocs – anondocs.aismarttalk.tech

open source · MIT

Open-source PII anonymiser (MIT) that detects and masks names, emails and identifiers **before a prompt leaves the network**. Runs on **self-hosted models via Ollama**, so sensitive documents never reach a third-party API, and ships as a public npm SDK ([@aismarttalk/anondocs-sdk](https://github.com/aismarttalk/anondocs-sdk)) that can be audited and deployed in place. **Cited by Wired Italia** among the ways to use AI without exposing sensitive data.

EDUCATION

Manager in Software Engineering (RIL) – accredited title, Master 1 level · CESI · Nancy – work-study programme	2016–2018
Bachelor – Unix Systems, Networks & Open Source (ASRALL) · IUT Charlemagne · Nancy	2015–2016
BTS IRIS – IT & Networks for Industry · Lycée Raymond Poincaré · Bar-le-Duc	2013–2015

LANGUAGES

English – Fluent – C1/C2, daily working language · **French** – Native – C2 · **German** – A2 and actively improving · **Luxembourgish** – Notions, learning