

Cayetano Soriano Gallego

AI in production, on platforms that don't break.

CTO and hands-on architect with 15+ years in tech, 8+ on AWS and 7+ as CTO. I've led two monolith-to-microservices transformations, and today I build LLM features, MCP servers and agentic engineering workflows that run in production.

hello@cayetano.dev · [linkedin.com/in/csorianogallego](https://www.linkedin.com/in/csorianogallego) · github.com/neoshadybeat · cayetano.dev · Madrid, Spain

15+7+299.99%

YEARS IN TECHYEARS AS CTOMONOLITH → MICROSERVICES MIGRATIONSPLATFORM UPTIME

What I build with AI

LLM features in real products

Claude generating bilingual listings and SEO copy for a marketplace; LLM symptom triage in a digital-health product; OpenAI and Gemini for chat and speech-to-text in a voice assistant. Tool calling and structured outputs, not free-form chat.

Claude APIOpenAIGemini

Tool callingSpeech-to-text

MCP servers & agent tooling

MCP servers that let agents read and act on product data safely, from a back-office admin server to a 7-module domain server (medication, schedule, calls, timeline...). Internal tools built on the Claude Agent SDK.

MCPClaude Agent SDKTypeScript

Python

Privacy, tests & cost as defaults

Personal data stripped before any prompt, feature-flag kill switches and switchable EU endpoints for GDPR; LLM calls covered by tests with a fake model transport; the right model per task to keep cost in check; Sentry monitoring like any other service.

GDPRPII redactionFeature flags

TestingSentry

Agentic engineering for teams

Claude Code across a dozen repos with project rules (CLAUDE.md) enforced in CI, AI agents in CI/CD and automated dependency updates. As CTO, I introduced LLM-assisted workflows into the engineering process.

Claude CodeCI/CDDependabotCursor

Experience

Independent

Apr 2026 – Present

Fractional CTO, Architect & AI Engineer

- Fractional CTO and hands-on architect/developer for early-stage companies: technical strategy, architecture reviews, cloud and delivery setup.
- Enolovers (client): cut app load time from ~10 s to under 1 s (resource loading, lazy vs eager data, DB indexes, Redis caching, rate limiting); Claude-powered bilingual content generation and an MCP admin server; CI/CD with AI agents, automated dependency updates and a test harness; security best practices; infrastructure in Terraform; Sentry, monitoring, analytics and per-call metrics.
- Own digital-health product (in development): LLM symptom triage with personal-data stripping, a voice companion with speech-to-text, and an MCP server with 7 tool modules.
- Deep-dive training in security, AI/LLM systems, software architecture and blockchain, applied directly to client work.

ShareTech

Mar 2023 – Apr 2026

Chief Technology Officer

- Owned the **technology vision and roadmap**; final say on cloud architecture, build-vs-buy and platform investments.
- Led cross-functional engineering squads; set standards for **microservices governance, reliability and developer experience**.
- Hands-on across **infrastructure automation (Terraform, ECS)**, AWS cost optimization and on-call/observability with Sentry + PagerDuty.
- **Vendor and partner contracts** (M2M and cybersecurity integrations with Telefónica); technical counterpart for Pamplona City Council.
- Operations & logistics tooling: **metrics, SLAs, reporting dashboards** and cost optimization.
- **End-to-end observability**: OpenTelemetry instrumentation, Datadog/New Relic, CloudWatch, Kibana, Sentry and PagerDuty on-call.
- Hands-on backend development; introduced **LLM-assisted workflows** into the engineering process.

Ride On

2019 – Mar 2023

Chief Technology Officer

- **Promoted from Architect to CTO** to own the full technology strategy of the eSharing mobility platform (remote since 2020).
- Scaled the engineering organization and introduced **OKRs** for the tech department.
- Drove **event-driven architecture on AWS Kinesis**; optimized container orchestration on **ECS**, cutting infrastructure cost ~30%.
- Directly accountable for **uptime (99.99%)**, **security and technical due diligence** with investors.
- **Vendor and partner contracts** (M2M and cybersecurity with Telefónica); technical counterpart for city councils in **Miami, Dundee, Leicester and Pamplona**.
- Operations & logistics: internal tooling, metrics, SLAs and reporting dashboards.
- **End-to-end observability**: instrumented every service with **OpenTracing / OpenTelemetry** (distributed tracing), structured metrics and alerting (New Relic, Datadog, CloudWatch, Kibana, Sentry, PagerDuty) — from request to database.

Ride On

2018 – 2019

Software Architect

- Led the complete platform re-architecture: **legacy Clojure monolith → 20+ containerized microservices on AWS ECS**.
- Authored the entire **Infrastructure-as-Code layer in Terraform**; automated CI/CD with Jenkins, images with Ansible + Packer.
- Established observability with Sentry and PagerDuty; removed the bottlenecks blocking **city-wide scale**.

Kimia

2017

Software Architect

- Designed and delivered **high-throughput backend services in Go**, backed by Redis and Elasticsearch.
- Built real-time monitoring and alerting pipelines that improved incident response times and reliability.

Packlink

2015 – 2017

Backend Engineer

- Key contributor to the **Packlink Pro and eBay shipping integration** backend.
- Drove the migration from a PHP monolith to **polyglot microservices (Java, Scala, Python)** with RabbitMQ, Cassandra and Redis.
- **Zero-downtime deployments** across high-traffic e-commerce operations.
- Platform **monitoring and on-call**: Kibana/Elasticsearch dashboards, CloudWatch, Sentry and PagerDuty for a high-availability product.

Skills & technologies

AI, agents & LLMs

- Claude / Anthropic API
- OpenAI API
- Gemini API
- MCP (Model Context Protocol)
- Claude Agent SDK
- AI agents
- Tool calling
- Structured outputs
- Prompt engineering
- Speech-to-text (Whisper)
- PII redaction / GDPR
- Claude Code
- Cursor
- GitHub Copilot
- Blockchain / Smart contracts

Cloud & infrastructure

- AWS
- Terraform
- ECS
- Serverless
- Kong
- Kinesis
- SNS / SQS

Backend

- Java / Spring Boot
- TypeScript / Node.js
- Python
- Scala
- PHP

Data & storage

- DynamoDB
- Elasticsearch
- MongoDB
- MySQL
- Redis / ElastiCache
- Cassandra
- Hadoop / Hive

DevOps, CI/CD & observability

- Docker
- Jenkins
- Ansible
- Packer
- Consul
- OpenTelemetry / OpenTracing
- Distributed tracing
- Datadog
- New Relic
- CloudWatch
- Kibana
- Sentry
- PagerDuty
- RabbitMQ
- Security hardening

Leadership

- Tech strategy & roadmap
- OKRs
- Hiring & team scaling
- Technical due diligence
- Cost management
- Security & incident response

Education & certifications

Universidad de Cádiz

Computer Systems Engineering
2007 – 2014

Megathon W8 Cádiz

Hackathon award

Claude Code in Action

Anthropic · 2026 · Verify
↗