

Cayetano Soriano Gallego

IA en producción, sobre plataformas que no se rompen.

CTO y arquitecto hands-on con más de 15 años en tech, 8+ en AWS y 7+ como CTO. He liderado dos transformaciones de monolito a microservicios, y hoy construyo funcionalidades con LLMs, servidores MCP y flujos de ingeniería agéntica que funcionan en producción.

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

15+

AÑOS EN TECH

7+

AÑOS COMO CTO

2

MIGRACIONES MONOLITO → MICROSERVICIOS

99.99%

UPTIME DE PLATAFORMA

Lo que construyo con IA

Funcionalidades LLM en productos reales

Claude generando fichas bilingües y textos SEO para un marketplace; triaje de síntomas con LLM en un producto de salud digital; OpenAI y Gemini para chat y voz a texto en un asistente de voz. Tool calling y salidas estructuradas, no chat libre.

Claude API

OpenAI

Gemini

Tool calling

Voz a texto

Servidores MCP y tooling para agentes

Servidores MCP que permiten a los agentes leer y actuar sobre datos de producto con seguridad, desde un servidor de administración del back office hasta uno de dominio con 7 módulos (medicación, agenda, llamadas, timeline...). Herramientas internas sobre Claude Agent SDK.

MCP

Claude Agent SDK

TypeScript

Python

Privacidad, tests y coste por defecto

Datos personales eliminados antes de cada prompt, kill switches con feature flags y endpoints UE conmutables por GDPR; llamadas a LLM cubiertas con tests sobre un transporte de modelo falso; el modelo adecuado para cada tarea para controlar el coste; monitorización con Sentry como cualquier otro servicio.

GDPR

Anonimización de datos

Feature flags

Testing

Sentry

Ingeniería agéntica para equipos

Claude Code en una docena de repos con reglas de proyecto (CLAUDE.md) verificadas en CI, agentes de IA en CI/CD y actualización automática de dependencias. Como CTO, introduje flujos asistidos por LLM en el proceso de ingeniería.

Claude Code

CI/CD

Dependabot

Cursor

Experiencia

o

Independiente

Abr 2026 – Actualidad

CTO fraccional, Arquitecto e Ingeniero de IA

- CTO fraccional y arquitecto/desarrollador hands-on para empresas en fase temprana: estrategia técnica, revisiones de arquitectura, setup cloud y de delivery.
- Enolovers (cliente): carga de la app de ~10 s a menos de 1 s (carga de recursos, datos lazy vs eager, índices en BD, caché Redis, rate limiting); generación de contenido bilingüe con Claude y un servidor MCP de administración; CI/CD con agentes de IA, actualización automática de dependencias y test harness; buenas prácticas de seguridad; infraestructura en Terraform; Sentry, monitorización, analíticas y métricas por llamada.
- Producto propio de salud digital (en desarrollo): triaje de síntomas con LLM eliminando datos personales, acompañante de voz con transcripción, y un servidor MCP con 7 módulos de herramientas.

- Formación intensiva en **seguridad, sistemas de IA/LLM, arquitectura de software y blockchain**, aplicada directamente en proyectos de clientes.

ShareTech

Mar 2023 – Apr 2026

Chief Technology Officer

- Responsable de la **visión tecnológica y el roadmap**; última palabra en arquitectura cloud, build-vs-buy e inversiones de plataforma.
- Lideré squads de ingeniería multifuncionales; definí estándares de **gobernanza de microservicios, fiabilidad y developer experience**.
- Hands-on en **automatización de infraestructura (Terraform, ECS)**, optimización de costes AWS y guardias/observabilidad con Sentry + PagerDuty.
- **Contratos con proveedores y partners** (integraciones M2M y ciberseguridad con Telefónica); interlocutor técnico con el Ayuntamiento de Pamplona.
- Herramientas de operaciones y logística: **métricas, SLAs, dashboards de reporting** y optimización de costes.
- **Observabilidad end-to-end**: instrumentación OpenTelemetry, Datadog/New Relic, CloudWatch, Kibana, Sentry y guardias con PagerDuty.
- Desarrollo backend hands-on; introduje **flujos asistidos por LLM** en el proceso de ingeniería.

Ride On

2019 – Mar 2023

Chief Technology Officer

- **Promoción de Arquitecto a CTO** para liderar toda la estrategia tecnológica de la plataforma de movilidad eSharing (en remoto desde 2020).
- Escalé la organización de ingeniería e introduje **OKRs** en el departamento técnico.
- Impulsé **arquitectura event-driven sobre AWS Kinesis**; optimicé la orquestación de contenedores en **ECS**, reduciendo el coste de infraestructura ~30%.
- Responsable directo de **uptime (99,99%), seguridad y due diligence técnica** con inversores.
- **Contratos con proveedores y partners** (M2M y ciberseguridad con Telefónica); interlocutor técnico con los ayuntamientos de **Miami, Dundee, Leicester y Pamplona**.
- Operaciones y logística: herramientas internas, métricas, SLAs y dashboards de reporting.
- **Observabilidad end-to-end**: instrumenté todos los servicios con **OpenTracing / OpenTelemetry** (distributed tracing), métricas estructuradas y alertas (New Relic, Datadog, CloudWatch, Kibana, Sentry, PagerDuty) — de la petición a la base de datos.

Ride On

2018 – 2019

Arquitecto de Software

- Lideré la re-arquitectura completa: **monolito Clojure legacy** → **20+ microservicios containerizados en AWS ECS**.
- Autor de toda la capa de **Infraestructura como Código en Terraform**; CI/CD automatizado con Jenkins, imágenes con Ansible + Packer.
- Observabilidad con Sentry y PagerDuty; eliminé los cuellos de botella que impedían la **escala a nivel de ciudad**.

Kimia

2017

Arquitecto de Software

- Diseño y entrega de **servicios backend de alto rendimiento en Go**, sobre Redis y Elasticsearch.
- Pipelines de monitorización y alertas en tiempo real que mejoraron los tiempos de respuesta a incidentes y la fiabilidad.

Packlink

2015 – 2017

Backend Engineer

- Contribuidor clave en el backend de **Packlink Pro** y la **integración de envíos de eBay**.
- Impulsé la migración de un monolito PHP a **microservicios políglotas (Java, Scala, Python)** con RabbitMQ, Cassandra y Redis.
- **Despliegues zero-downtime** en operaciones e-commerce de alto tráfico.

- **Monitorización y guardias** de la plataforma: dashboards Kibana/Elasticsearch, CloudWatch, Sentry y PagerDuty para un producto de alta disponibilidad.

Tecnologías y habilidades

IA, agentes y 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 e infraestructura

AWS Terraform ECS

Serverless Kong

Kinesis SNS / SQS

Backend

Java / Spring Boot

TypeScript / Node.js

Python Scala PHP

Datos y almacenamiento

DynamoDB Elasticsearch

MongoDB MySQL

Redis / ElastiCache

Cassandra Hadoop / Hive

DevOps, CI/CD y observabilidad

Docker Jenkins Ansible

Packer Consul

OpenTelemetry / OpenTracing

Distributed tracing

Datadog New Relic

CloudWatch Kibana

Sentry PagerDuty

RabbitMQ

Security hardening

Liderazgo

Estrategia tecnológica y roadmap

OKRs

Contratación y escalado de equipos

Due diligence técnica

Gestión de costes

Seguridad y respuesta a incidentes

Formación y certificaciones

Universidad de Cádiz

Ingeniería Técnica en Informática de Sistemas

2007 – 2014

Megathon W8 Cádiz

Premio hackathon

Claude Code in Action

Anthropic · 2026 · [Verificar ↗](#)