

ADRIÁN VÁZQUEZ

Ingeniero de IA | RAG • Agentes LLM • Servidores MCP

Barcelona, España (disponible en remoto y reubicación) | +34 680 325 594 | adrian@vazquezdev.pro
linkedin.com/in/adrivaz | github.com/avaazquezz | vazquezdev.pro

RESUMEN PROFESIONAL

Ingeniero de IA especializado en sistemas RAG (Retrieval-Augmented Generation) en producción, agentes LLM y servidores MCP (Model Context Protocol). Único ingeniero de IA en CSR Online, responsable de 3 sistemas de IA en producción de principio a fin. Despliegue con APIs de LLM en la nube o modelos autoalojados cuando los datos no pueden salir de la empresa, con salidas validadas, evals, aprobación humana y observabilidad de LLM.

COMPETENCIAS TÉCNICAS

- IA y LLM: RAG, agentes LLM, orquestación multiagente, tool calling, salidas estructuradas, prompt engineering, evaluación de LLM (evals), MCP (Model Context Protocol), LangGraph, LangChain, Claude, OpenAI, Gemini, Groq, Ollama
- Recuperación y datos: Qdrant, ChromaDB, pgvector, bases de datos vectoriales, embeddings, búsqueda híbrida (RRF/DBSF), búsqueda semántica, PostgreSQL, Redis, SQL Server (T-SQL)
- Backend y DevOps: Python, FastAPI, Pydantic, Node.js, C#/.NET, APIs REST, Docker, Docker Compose, Traefik, CI/CD, Linux, Prometheus, Langfuse
- Frontend y automatización: JavaScript, TypeScript, Vue 3, Nuxt, Vuetify, Tailwind CSS, n8n
- Idiomas: español (nativo), catalán (nativo), inglés (competencia profesional, B2)

EXPERIENCIA PROFESIONAL

Ingeniero de IA — CSR Online

Barcelona (remoto) · Mar 2026 – Actualidad

- Único ingeniero del departamento de IA, dependiendo directamente del cofundador y CTO; responsable de principio a fin de 3 sistemas de IA en producción (infraestructura, base de datos vectorial, backend, despliegue), incluidos sistemas multiagente y arquitecturas RAG empresariales autoalojadas.
- Desarrollé Enterprise Context AI, un asistente híbrido RAG + SQL autoalojado (Python, Qdrant, Docker): la dirección consulta el correo, los documentos internos y la base de datos operativa desde una única interfaz conversacional, todo on-premise.
- Ingesta de correo en tiempo real con Exchange Journal más archivos PST históricos; herramienta SQL invocable por el LLM para respuestas precisas cuando la búsqueda vectorial no basta.

Ingeniero Full Stack y de IA independiente — VazquezDev

Remoto · Sep 2023 – Actualidad

- Diseñé, desarrollé y desplegué de principio a fin 6 SaaS de gestión de negocio para familiares, amigos y conocidos, desde la arquitectura y la base de datos hasta producción (sep 2023 – sep 2025).
- Especializado en IA desde oct 2025: sistemas RAG y flujos multiagente con responsabilidad total, de la arquitectura al despliegue, en paralelo a mi puesto en CSR Online (ver Proyectos; vazquezdev.pro).

Ingeniero de backend e infraestructura de datos — Senyum SL

Prácticas · Sep 2024 – Jul 2025

- Administración de bases de datos SQL Server y T-SQL avanzado (consultas, vistas, índices, procedimientos almacenados) con ajuste de rendimiento.
- Automatizaciones e integraciones de datos en C#/.NET (a3ERP), modelado de datos y cuadros de mando en Power BI; mantenimiento de la infraestructura backend para alta disponibilidad.

PROYECTOS

- ReportAI (reportai.vazquezlabs.com): agente LangGraph (FastAPI, Claude, Groq Whisper, PostgreSQL, Docker) que convierte una nota de voz de 90 segundos en un informe terminado en minutos, con la plantilla Word de cada cliente y sin transcribir ni maquetar a mano. Extracción validada por esquema, suite de evals propia y aprobación humana; 10–15 informes/mes en producción.
- Qdrant-MCP (github.com/avaazquezz/Qdrant-MCP): servidor MCP de código abierto (MIT) que expone 49 herramientas de Qdrant a cualquier cliente MCP, con búsqueda híbrida (RRF/DBSF), modo de solo lectura, protección SSRF y métricas Prometheus; en PyPI y Docker.

FORMACIÓN ACADÉMICA

- Técnico Superior en Desarrollo de Aplicaciones Multiplataforma, Institut Pedralbes
- Técnico en Sistemas Microinformáticos y Redes, Institut Pedralbes