

Antonio Puerto Borreguero

AI Engineer · Soluciones de IA de extremo a extremo — del LLM al firmware

Sevilla, España · Híbrido / Remoto · info@antoniopuerto.com · 665 694 029

github.com/PuertOcho · linkedin.com/in/antonio-puerto-borreguero · antoniopuerto.com



PERFIL

Ingeniero informático que lleva la IA a producción de extremo a extremo: del modelo de lenguaje y los agentes a los microservicios que los sirven y el hardware donde corren. Trayectoria en IA desde 2020, evolucionando con ella de forma autodidacta: de los fundamentos académicos (algoritmos genéticos, MCTS, árboles de decisión) a **tres generaciones de asistentes propios** — la última, **tony**, una plataforma agéntica con más de 1.000 commits y 36 releases — cuyas técnicas (agentes, MCP, observabilidad) aplico directamente en mi trabajo profesional.

Mi diferenciador: vengo de los sistemas empotrados (Renesas, STM32, FreeRTOS, CAN), así que no me quedo en el prototipo en la nube — entiendo el dispositivo físico de punta a punta. Hoy diseño e implemento mejoras de IA en flujos de trabajo industriales y construyo agentes, integraciones **MCP** y arquitecturas de microservicios que funcionan en producción.

STACK TÉCNICO

IA & Agentes	LLMs · IA generativa · Agentes · RAG · MCP · function calling · LiteLLM · Langfuse · LangChain/LangGraph/LangSmith · NLU · STT/TTS · OCR (Tesseract) · prompt engineering · IA clásica (genéticos, MCTS, ID3/C4.5)
Backend & Cloud	Java · Spring Boot · C#/NET (WPF) · microservicios · Eureka · REST · arquitectura hexagonal · PostgreSQL · MariaDB · Redis · Kafka · Maven · JUnit · Jenkins
Embedded & HW	C · C++ · Renesas RX · STM32 · ESP32 · FreeRTOS · CAN · UART · SPI · I2C · protocolos industriales
DevOps	Docker · Docker Compose · Nginx · Grafana · Prometheus · Linux · self-hosting · Git
Frontend	Vue.js / Quasar · TypeScript · Svelte · Ionic

EXPERIENCIA

Ingeniero de Firmware e IA · I+D

abr. 2023 – actualidad · Sevilla

COBER (Control of Biomedical Embedded Robotics)

I+D en robótica biomédica y sistemas empotrados, con responsabilidad creciente en aplicar IA al producto y al flujo de trabajo — apoyada en la ingeniería de agentes de mi I+D personal.

- Diseñé y desarrollé la **plataforma interna de gestión de placas hardware y contratos**: Spring Boot 3 + MariaDB + JWT, cliente WPF (.NET), escáner de códigos de barras, Docker y CI/CD con Jenkins — 7 módulos y +40 endpoints REST (310 commits).
- Incorporé **clasificación automática de documentos y contratos con OCR e IA** (Python/Tesseract): extracción de campos, clasificación por fabricante con confianza y validación cruzada.
- **Colaboración con MP Lifts (desde sept. 2024)**: migración completa del protocolo propietario sobre CAN (v2 a v3) en el controlador de ascensores y su interfaz web de diagnóstico — **9/9 módulos verificados en hardware real**, 12 servicios de monitorización, +200 commits entre firmware (C/C++, FreeRTOS, Renesas RX) y frontend (Vue.js/Quasar, i18n en 6 idiomas).
- Reduje el tiempo de respuesta de la petición de eventos **de 12 s a 0,7 s (-94 %)** optimizando timeouts y reintentos del stack de comunicaciones.
- Automaticé la verificación end-to-end con una suite (PowerShell + Playwright) que cubre **+130 endpoints HTTP** contra hardware real.
- Construí **servidores MCP** que conectan agentes LLM con herramientas reales de ingeniería: IDE e2 Studio (10 tools; extensión **publicada en Visual Studio Marketplace**) y bus CAN industrial (17 tools, chat LLM con function calling).

Programador Full Stack

sept. 2021 – abr. 2023

Insinno España · prácticas extracurriculares mar.–sept. 2021

- Desarrollo de la **aplicación de gestión de productos químicos para BASF** (la mayor empresa química del mundo): backend Java/Spring Boot con microservicios, APIs REST, PostgreSQL/Redis y mensajería con Kafka.
- Producto escalable bajo estándares de calidad de cliente multinacional: arquitectura hexagonal, testing con JUnit y CI sobre Maven.

Vendedor (equipo de Montaña)

sept. 2020 – abr. 2021

Decathlon · Camas, Sevilla

- Atención al cliente y asesoramiento técnico, compaginado con los estudios.

PROYECTOS DESTACADOS (I+D PROPIA EN IA)

tony — Plataforma agéntica de 3ª generación (2025–act.)

Multi-proveedor (OpenAI, Anthropic, Google vía LiteLLM, fallback local Ollama) con MoE propio: votación paralela de modelos con consenso y debate. Cerebro bio-inspirado sobre LangGraph, +20 servicios MCP, cliente de escritorio (Tauri + Svelte), evaluación semántica continua (122 casos con juez LLM), observabilidad (Langfuse, Grafana, Prometheus) y hardware propio (ESP32-S3). +1.000 commits, 36 releases, +6.300 tests.

puertocho-assistant — Asistente de voz (2ª generación, 2025)

Microservicios Spring Boot + Eureka con capa de voz E2E: Whisper local, gestor de intenciones LLM-RAG con votación MoE, descomposición dinámica de tareas y cadena TTS (Azure, Kokoro, XTTS-v2, F5/E2). Desplegado sobre Raspberry Pi.

nuka — Asistente (1ª generación, 2023–2024)

El origen de la serie, en la primera ola de la IA generativa: GPT + Whisper + Azure TTS + DALL-E + RAG sobre Obsidian. App Ionic/Angular + servidor Spring Boot. 485 commits.

Servidores MCP a medida

e2studio-mcp (build/debug de firmware Renesas; extensión en Visual Studio Marketplace), mp-can-mcp (bus CAN industrial + chat LLM), taiga-mcp-ms (gestión de proyectos).

FORMACIÓN

Grado en Ingeniería Informática — Universidad de Sevilla · 2016 – 2021 · Fundamentos de IA (2020): algoritmos genéticos (TSP), MCTS, árboles de decisión (ID3/C4.5)

Inteligencia Artificial Aplicada a la Empresa — EDUCATIC GAP PUE · nov. 2022 – feb. 2023

Arquitectura Big Data — Fundación CONFEMETA · jun. 2021 – sept. 2022

IDIOMAS

Español (nativo) · Inglés (B2 — nivel medio-alto, uso profesional)