

PABLO CÁCERES RAMOS

AI Engineer | IA generativa | Sistemas multiagente | MLOps

pablocaceres1999@gmail.com | LinkedIn (9 recomendaciones) | pcaceresr.es

PERFIL

AI Engineer con más de 5 años de experiencia en IA y desarrollo de software, especializado en sistemas de IA generativa, RAG y arquitecturas multiagente en producción. Trayectoria end-to-end sobre AWS y Azure, desde el diseño y el desarrollo hasta el despliegue, la evaluación y la observabilidad. Experiencia en soluciones para Verisure, Telefónica, Enel X North America, Iberdrola y proyectos de I+D de la Unión Europea.

COMPETENCIAS CLAVE

IA generativa: LLMs, RAG, AI Agents, Multi-Agent Systems, MCP, A2A, LangChain, LangGraph, CrewAI, Embeddings, Vector Search, Text-to-SQL, Text-to-API, Text-to-NoSQL, Document Intelligence, LLM-as-judge, Pydantic

Cloud / MLOps: AWS (Bedrock AgentCore, Lambda, CloudWatch), Azure (OpenAI, AI Search, AI Foundry, Container Apps), Docker, Terraform, CDK, pytest, OpenTelemetry, Langfuse, Phoenix

Backend y datos: Python, FastAPI, REST APIs, PostgreSQL, MongoDB, DynamoDB, Snowflake, StarRocks

ML / DS: Scikit-learn, Keras, RNN, LSTM, GRU, CNN, XGBoost, CatBoost, Pyomo, NumPy, Pandas

FORMACIÓN IMPARTIDA Y DIVULGACIÓN TÉCNICA

- Diseño e **impartición de formaciones técnicas** sobre construcción end-to-end de sistemas RAG y chatbots: integración con LLMs, frontend, gestión de sesiones, memoria conversacional y memoria a largo plazo.
- Impartición de cursos de introducción a la IA generativa para perfiles no técnicos: fundamentos, prompting, gestión del contexto y de los tokens, y aplicación de la IA a diferentes tareas usando n8n.

EXPERIENCIA PROFESIONAL

AI Engineer | IA generativa y sistemas multiagente

abr. 2026 - actualidad

Verisure | Remoto

- Diseño, desarrollo y despliegue sistemas multiagente sobre AWS Bedrock AgentCore, integrados mediante A2A y apoyados en Lambda y DynamoDB.
- Creé **desde cero** una librería interna que estandariza la creación de agentes y sistemas multiagente.
- Desarrollo agentes con herramientas expuestas mediante MCP, memoria a corto y largo plazo y observabilidad con OpenTelemetry, CloudWatch y Langfuse.

AI Engineer | IA generativa y sistemas multiagente

ago. 2024 - abr. 2026

Telefónica TID | Remoto

- Diseñé y desarrollé el end-to-end de agentes especializados Text-to-SQL, Text-to-NoSQL y Text-to-API, integrados en una arquitectura multiagente desplegada en producción sobre Azure.
- Implementé un RAG corporativo multifuente y multimodal con respuestas citadas, reindexación automática y evaluación mediante LLM-as-judge para monitorizar la calidad de las respuestas.
- Desarrollé soluciones end-to-end (backend, frontend, bases de datos y APIs) y diseñé arquitecturas Azure para bots y aplicaciones de Microsoft Teams conforme a los requisitos de seguridad de CISO.
- Desarrollé el núcleo de una herramienta de IA **premiada en el hackathon interno Telefónica Equinox 2025**.

Data Scientist | AI Engineer

oct. 2022 - ago. 2024

Ayesa | Enel X North America, Acciona, Iberdrola y proyectos de I+D de la Unión Europea

- Participé en el ciclo completo de proyectos de I+D de la Unión Europea (requisitos, coordinación con socios internacionales y desarrollo de algoritmos de IA) y propuse estrategias de IA generativa con RAG, Text-to-API, LangChain y fine-tuning.
- Desarrollé estrategias de Machine Learning para multinacionales del sector energético, con cálculo de KPIs en tiempo real y procesamiento ETL de datos para proyectos de Deep Learning.
- Formé parte del **único equipo que superó una licitación técnica de Iberdrola**, desarrollando modelos RNN, LSTM, GRU y CNN con Keras.
- Reduje la ejecución de un sistema de IA de **más de 300 h a menos de 10 h (más de 30 veces)** mediante el rediseño de estructuras de datos y optimizaciones internas.

Software Engineer | AI Engineer

abr. 2022 - oct. 2022

Solartree | Remoto

- Mantenimiento web y asesoramiento tecnológico; desarrollé una PoC de chatbot con la API de GPT-3 en Python antes del lanzamiento de ChatGPT.

AI Engineer

ene. 2020 - nov. 2020

Universidad de Huelva | Huelva

- Desarrollé una plataforma end-to-end para la selección de turbinas hidráulicas: enfoques de IA para clasificación, frontend, backend y base de datos, con metodología ágil.

PROYECTO DESTACADO

- **Clasificación multimodal (NLP):** alcancé la **4.ª mejor puntuación** en una competición CodaLab de clasificación de perfiles de Twitter con Transformers.

EDUCACIÓN

Postgrado en Big Data [Universidad de Huelva](#)

2021 - 2023

TFM: Matrícula de Honor.

Grado en Ingeniería Informática [Universidad de Huelva](#)

2017 - 2021

Premio al Mejor TFG sobre Fuentes Renovables.

Comunicación y Oratoria [Escuela de Neurocomunicación y Liderazgo Humanista](#)

2024 - 2025

Duración de 111 h

CERTIFICACIONES Y FORMACIÓN

- Microsoft Azure Fundamentals (AZ-900).
- MongoDB - 4 certificados:
Building AI Agents, AI-Powered Search, Schema Design Patterns & Anti-patterns; Advanced Schema Patterns.
- LangChain: Deep Research Agents.

IDIOMAS

Español: nativo.

Inglés: B2 certificado · Uso profesional en proyectos internacionales de I+D de la Unión Europea y para Enel X North America.