

**Javier Rodríguez Fontiveros**  
**Data Scientist**  
**Madrid, Spain**  
**+34 635 893 405**

[Correo de contacto](#)  
[Perfil de LinkedIn](#)  
[Perfil de Kaggle](#)  
[Portfolio Github](#)

## Resumen

Data Scientist con sólido trasfondo en ingeniería y expertise en el ciclo de vida de datos **E2E (End-to-End)**. Especialista en la construcción de modelos de Machine Learning, flujos **MLOps/LLMOps** y generación de Business Intelligence. Experto en **Python y SQL**, con experiencia demostrada integrando ecosistemas **Open Source y GenAI** para desplegar soluciones de IA escalables, auditables y orientadas a la toma de decisiones estratégicas en entornos de alta complejidad.

## Experiencia profesional

**IMDEA Networks Institute | Marzo 2025 - Enero 2026**

**Data Engineer (Contratado tras periodo de prácticas) | Diciembre 2025 - Enero 2026**

- **Estrategia de Datos:** Automatización de pipelines mediante **GitLab CI/CD**, reduciendo errores manuales y garantizando ejecución 24/7 mediante el despliegue de runner grupal en nube privada.
- **Arquitectura de Datos:** Diseño de flujos **ETL (SQL Server a PostgreSQL)** y optimización de esquemas para el Data Warehouse, incluyendo un sistema de sincronización histórica diaria para auditorías.
- **Ingeniería de IA:** Implementación de flujos **OCR (PyTesseract, Camelot)** para la detección automática de inconsistencias financieras y mejorando la trazabilidad de informes BI.
- **Liderazgo Técnico:** Punto de contacto de **MLOps** para consultoría externa, asegurando estándares de despliegue y calidad de código.

**Data Scientist & MLOps Intern | Marzo 2025 - Diciembre 2025**

- **Procesamiento de Lenguaje Natural (NLP):** Clasificación de texto (**22 subcategorías de hacienda**) con 91% de efectividad mediante indexación vectorial FAISS + K-NN, Random Forest sobre **embeddings** y análisis de similitud semántica.
- **Automatización & Scraping:** Sistema de web scraping avanzado con **Selenium y ScrapeGraphAI** (LLM-Based Qwen), integrando pipelines en Airflow.
- **Experimentación ML:** Gestión del ciclo de vida en modelos con **MLflow**, realizando EDA avanzado y Process Mining (pm4py) para la detección de cuellos de botella.
- **Despliegue de IA:** Creación de prototipos interactivos en **Streamlit** bajo contenedores **Docker** e implementación de sistemas RAG con Mistral Instruct, ChromaDB y wrapper OpenAI como agente.
- **GPU Computing:** Conexiones SSH al datacenter local usando **NVIDIA RAPIDS** para procesamiento acelerado mediante servidores **Linux**.
- **Proyectos clave:** Desarrollo de producto (Report Generator) y **cumplimiento normativo con análisis de datos** (Compliance) y apoyo en la evaluación técnica de licitaciones públicas de alto valor.

**Agencia de Medio Ambiente | Abril 2022 - Diciembre 2022**

**Prácticas de Ingeniería Química**

- Optimización de procesos en plantas de tratamiento de aguas (**humedales aireados**) y **cálculo de prognosis** para la planificación de capacidad y evaluación de diseños alternativos.

## Habilidades Técnicas

- **Lenguajes:** Python, SQL, Java (Spring Boot), Bash.
- **Machine Learning:** Scikit-learn, XGBoost, PyTorch, Keras, TensorFlow, SHAP, Optuna.
- **Generative AI & NLP:** LLMs (Qwen, Mistral), RAG (LangChain, Langfuse), Vector DBs (ChromaDB, FAISS), MCP, Hugging Face Transformers, BERT, ScrapeGraphAI.
- **Data Engineering:** PostgreSQL, SQL Server, Redis, MinIO, MongoDB, PySpark, ETL Pipelines, Process Mining (pm4py), Pandas, NumPy, Web Scraping (Selenium, BeautifulSoup).
- **MLOps & Orquestación:** MLflow, Apache Airflow, Docker, GitLab CI/CD, Prometheus, Grafana.
- **Cloud (AWS):** S3, Lambda, RDS, SQS/SNS, Terraform, LocalStack.
- **Visualización & APIs:** Power BI, Tableau, Streamlit, FastAPI, Flask, Uvicorn, Plotly, Seaborn, Matplotlib.
- **Entornos:** RAPIDS (GPU computing), Linux, poetry, anaconda.

## Proyectos

- **Sistema automatizado de trading en IBKR:** levantamiento de servicios de MLflow, Airflow, Minio y postgres. Implementación de lógica de compra/venta y acceso a la api de TWS mediante un bot programado.
- **ChatBot vih360 (GenAI & Backend) | Python & AWS:** Co-desarrollo de un chatbot para la FELGTBI+ integrando Google Gemini y Google Places API. Implementé la arquitectura de backend en AWS y un panel de analítica sociodemográfica en tiempo real.
- **Blockchain Simulation | Java:** Desarrollo de una red de blockchain funcional con gestión de wallets, transacciones firmadas digitalmente y simulación de smart contracts.

## Premios y Logros

- **JenPlane's Datathon - DataBri-X feat IMDEA Networks (3er puesto):** Pipeline escalable optimizando el reporting de datos de red.
- **Schneider Electric Hackathon x Nuwe (Top 15 de 335):** Modelo predictivo de contaminantes atmosféricos y detección de anomalías en sensores industriales.
- **Caixabank Tech Hackathons x Nuwe (Rank 11/81 python & 35/261 Java):** Comunicación entre microservicios y desarrollo de sistemas de subasta en tiempo real.

## Educación y Certificaciones

- **Data Science Bootcamp,** The Bridge | Madrid (2024 - 2025)
- **Grado en Ingeniería Química,** Universidad de Sevilla (2017 - 2023)
- **Certificaciones:** AI & LLM Engineering Mastery y MCP:Tools for LLMs (O'Reilly), Kaggle Learn / AWS Educate (+ 100 h cursos online)

## Idiomas y Soft Skills

- **Idiomas:** Español (nativo), Inglés (Competencia Profesional / B2 - C1)
- **Soft Skills:** Data Storytelling y comunicación con stakeholders, resolución innovadora de problemas, colaboración técnica en equipos multidisciplinares y adaptabilidad tecnológica/aprendizaje rápido..