

Guillermo Quintana Pelayo

AI Engineer

Senior AI Engineer specialized in production LLM/RAG systems on GCP, with +5 years of experience delivering end-to-end AI products from real-time data pipelines to APIs.

✉ gqpsantander@gmail.com

☎ +34 672 027 380

🌐 [guilleqp](https://www.linkedin.com/in/guilleqp)

🔗 [GuilleQP](https://github.com/GuilleQP)

WORK EXPERIENCE

AI Engineer

dealroom.co

📅 Jan 2022 - current 📍 Netherlands - Spain

- Owned production LLM systems for enrichment, classification, and generation end-to-end.
- Shipped RAG + hybrid retrieval with chunking and reranking to improve grounding and relevance.
- Built semantic search end-to-end: embeddings, vector index, retrieval, backend, and UI integration. With a latency <5s in a 5M records pool.
- Operated ML services on GCP (containerized deployments, CI/CD, logging, alerting).
- Led small teams delivering AI features to production.
- Built real-time ingestion and data APIs on GCP for scalable, low-latency pipelines.
- Designed evaluation and monitoring for LLM/NLP systems (benchmarks, regression tests, human review loops).
- Delivered full-stack projects: backend services, data processing, ML inference, and frontend integration.

Data Scientist

dealroom.co

📅 Aug 2020 - Dec 2021 📍 Amsterdam, NL

- Trained RoBERTa-based classifiers for multi-label text classification with general accuracy over 85%.
- Trained NER models for entity extraction from unstructured text.
- Productionized data/ML workflows with reusable pipelines and metrics.
- Built automation pipelines to process and validate large datasets.
- Ran error analysis and iterative model improvements to increase accuracy.

EDUCATION

Computer Science
Engineering

[University of Cantabria,
Spain](https://www.univ Cantabria.es)

📅 2015 - 2019

Master's degree,
Artificial Intelligence

[Maastricht University,
Netherlands](https://www.maastrichtuniversity.nl)

📅 2019 - 2022

SKILLS

- **Cloud & MLOps**
GCP (Vertex AI, Cloud Run, Pub/Sub, BigQuery), Docker, CI/CD, LLMOps, Terraform
- **AI & Machine Learning**
LLMs, Hybrid RAG, Semantic Search, Chunking, Reranking, Transformers
- **Languages & Frameworks**
Python, FastAPI, Hugging Face, PyTorch, LangChain

LANGUAGES

- Native Spanish
- C1 English
- N5 Japanese