

Guillermo Hoyo Bravo

GENAI ENGINEER — MADRID, SPAIN

guillehoyob@gmail.com · github.com/guillehoyob · linkedin.com/in/guillermo-bravo

PROFESSIONAL SUMMARY

GenAI engineer with around two years building production AI systems. I design, build, evaluate, and lead RAG architecture and multi-agent orchestration — from retrieval pipelines and ingestion libraries to evaluation suites that gate every deploy. I work AI-pair-engineering as a documented method: I direct the design and review the build commit by commit, so the human owns the decisions and the AI accelerates the work.

EXPERIENCE

AI Engineer — IDEA

[PLACEHOLDER – dates to confirm]

Sole AI engineer at IDEA: designed, built, and leads the company's production RAG platform.

- Designed, built, and now lead a production RAG platform on Docker, Azure CI/CD, PostgreSQL, and Qdrant, with an endpoint factory and reusable ingestion/retrieval libraries so new assistants ship from shared components.
- Integrated a RAGAS evaluation module and Langfuse observability so retrieval and answer quality are measured, not assumed.
- Shipped two assistants to production — an HR assistant and an electrical-engineering-rules assistant — with strong measured results
[PLACEHOLDER – metrics to confirm].
- Mentor an engineer building a LangGraph agent on the platform.

RAG Chatbot Lead (AI-assisted) — Zelebrix

[PLACEHOLDER – ongoing, to confirm]

Joined to lead one AI feature — a RAG chatbot for a multi-tenant payments SaaS.

Built and validated; scheduled to deploy to production in

[PLACEHOLDER – target window, ~a few months]. Ongoing collaboration, role still being defined (currently unpaid).

- Designed and led the build of the RAG chatbot, AI-assisted and reviewed commit by commit.
- Built hybrid retrieval: dense pgvector plus Spanish Postgres full-text search, fused with Reciprocal Rank Fusion, then a cross-encoder reranker and a relevance gate.
- Designed privacy-by-design: PII scrubbing before the LLM, and data queries executed locally so the external LLM never sees tenant data — a boundary that holds by construction.
- Built a custom evaluation suite — a 150-question golden set, a retrieval probe, and a RAGAS-lite scorer — and an LLM benchmark that showed with data the bottleneck was retrieval (synonyms), not the generation model.

- Extended a production multi-agent system on Semantic Kernel; added plugins for Text-to-SQL (Databricks), document extraction (Azure Blob Storage), translation, SharePoint lookup, and a Service Desk SOAP service in under four weeks.
- Built a Text-to-SQL agent with an enriched metadata layer and a custom evaluation framework with a golden dataset — 100% accuracy on controlled tests before every deploy. Implemented RBAC impersonation so every query runs under the requesting user’s identity.
- Built FastAPI endpoints with async token streaming, Pydantic-validated I/O, selective per-plugin memory (MongoDB), and end-to-end impersonation; deployed via Azure DevOps CI/CD to AKS.
- Contributed to a hybrid RAG pipeline on Azure AI Search (semantic + BM25 + cross-encoder reranking) with permission-aware retrieval; reviewed intern PRs and set plugin contribution criteria.

- Automated data cleaning and processing for large datasets with Python and KNIME; researched OCR digitisation (Pytesseract, OpenCV, Azure AI Studio).

EDUCATION

- **MSc in Data Science** — Autonomous University of Madrid (UAM), 2021–2023. GPA 7.63/10. Time-series forecasting, deep learning, NLP, optimization.
- **BSc in Computer Engineering** — UAM, 2016–2021. Erasmus at the University of Bergen, Norway. GPA 7.03/10.

CERTIFICATIONS

- Hugging Face — “Fine-tuning Language Models” (Oct 2025)
- AlgoExpert — “Machine Learning Expert” (Sep 2025)

LANGUAGES

- Spanish — native
- English — B2; bilingual education, Erasmus in Norway
- Vietnamese — learning

SKILLS

GenAI & RAG: Semantic Kernel, Azure OpenAI, hybrid RAG (vector + BM25 + reranking), multi-agent orchestration, prompt engineering, RAGAS evaluation, cross-encoder rerankers, Reciprocal Rank Fusion, LangGraph, Langfuse.

Cloud & Data: Azure (AI Search, Blob Storage, Key Vault, Document Intelligence, AKS, DevOps), Databricks, Docker, CI/CD, PostgreSQL/pgvector, Qdrant.

Programming: Python (FastAPI, Pandas, PySpark, Scikit-learn), SQL, Power BI.