Juan Pablo Osorio Rivera

jp.osorio4102@gmail.com | +34 612 255 190 | Barcelona, Spain | linkedin.com/in/jp-or | github.com/garasnote

Profile

Data scientist and AI engineer who builds retrieval systems together with the measurement that keeps them honest: a corpus, a retriever, an evaluation set, and results that include the configurations that lost.

Work Experience

Data Analyst (Profesional Universitario) Apr 2024 – Aug 2024

Gobernación del Meta — Department of Health · Villavicencio, Colombia

- Cleaned and validated epidemiological records from SIVIGILA, Colombia's national public-health surveillance system, and mapped poisoning cases across the Meta department so a small team could target inspections instead of spreading them evenly.
- Automated the consolidation of the regulated chemical-retailer registry with Google Sheets and Apps Script, removing manual entry errors and turning a paper list into a municipality-level map used to allocate field resources.

Projects

HireSignal: a Job-Market Warehouse https://garasnote.github.io/hiresignal/

Python · SQL · DuckDB · pytest · GitHub Actions · dbt

- A warehouse of 233,000 real job postings harvested from eleven applicant-tracking platforms: conformed and deduplicated in SQL, modelled into a DuckDB star schema with seven marts, each answering one business question.
- The dashboard is generated from those marts - semantic HTML, one page per daily snapshot, live on GitHub Pages - behind 17 tests and a GitHub Actions gate, with a daily 1,200-board panel snapshot.

Retrieval Service with a Measured Evaluation 2026

BM25 from scratch · MiniLM ONNX · reciprocal-rank fusion · FastAPI

- Retrieval service with BM25 implemented from scratch, MiniLM dense retrieval and reciprocal-rank fusion, compared on 15 hand-written questions: hit@1 0.80, MRR 0.85 for BM25, with every configuration published in the README instead of only the best one.

Multi-Modal ML for Snow Coverage Estimation 2025 – 2026

PyTorch · Hugging Face · Lightning · OpenCV · Optuna · SHAP · SLURM / HPC

- Paper under review (first author): *Multi-Modal Machine Learning for Image-Based Snow Coverage Estimation*, Computers & Geosciences. **Best Presentation Award**, University of Ljubljana, 2026.
- Built a DINOv3-based segmentation model estimating per-pixel snow presence from webcam imagery, reaching **0.90 mIoU** across 12 weather stations and replacing manual on-site observation for the Slovenian Meteorological Agency.
- Fused visual features with meteorological data, beating the image-only baseline by **+0.10 AUROC**: Optuna's TPE for tuning, SHAP to attribute the predictions.

Education

MSc Artificial Intelligence Universitat Pompeu Fabra / University of Ljubljana Sep 2024 – Aug 2026

€20,000 Erasmus Mundus scholarship · 3% admission rate · thesis graded 10/10

BSc Chemical Engineering Universidad de Los Andes, Bogotá · full merit scholarship Jan 2019 – Jun 2023

Skills

Programming: Python, R, SQL, MATLAB, Bash

GenAI & Web: RAG pipelines, prompt design, simple agents, Ollama, pgvector / PostgreSQL, FastAPI, Flask, REST APIs

Data & Infra: Docker / Compose, Linux, Git, GitHub Actions, YAML, SLURM, dbt, DuckDB, Parquet, Polars, Dask

Classical ML: Scikit-learn, LightGBM, XGBoost, OpenCV, SciPy, hmmlearn, Optuna, SHAP

BI & Reporting: SQL, Excel (advanced), Google Sheets / Apps Script, Metabase, Power BI (DAX measures), Streamlit

Languages: Spanish (native) · English (C2, TOEFL)