EDUARDO GONZÁLEZ-MESONES

Data and AI Engineer | Banking, pharma and regulated sectors

Madrid, Spain (open to remote) | e.gzlzmesones@gmail.com

edu.glemesfft.site | linkedin.com/in/edugonzalezmesones | github.com/edumesones

PROFILE

AI engineer with six years of experience building machine learning and LLM systems that run in production. I spent four years in data science at regulated banks, owning models from raw data to a served endpoint, and since 2024 I have focused on agents, retrieval and workflow automation. As founder of an independent AI consultancy I take client problems from a loose brief to a deployed system: I agree with the business owner on how success will be measured, build the evaluation before iterating on prompts or models, and keep a person in the loop wherever a mistake is costly. AI coding agents are part of my daily work, with my own review and testing routine on top of what they generate. I work in English with international clients and distributed teams.

SELECTED SYSTEMS

Competitive intelligence agent | Listed pharma company | AWS Bedrock AgentCore, Strands, MCP

Streamed agent behind an MCP server. A guard with no tools and no memory classifies each request before the main agent runs, and rejects by default if the classifier fails. Three internal sources are queried concurrently, so latency depends on the slowest source only. User identity is taken from a signed token claim, which keeps every user's memory isolated from the others.

Legal RAG with verifiable citations | Own project | Python, pgvector

The retriever returns an article number plus the exact passage, so every answer can be traced to its source. I built the evaluation harness first; it exposed a groundedness metric that was dividing by retrieved chunks when it should divide by claims in the answer. Citation recall 0.70, precision 0.575.

AML alert scoring ensemble | Regulated banking | PySpark, GNN, gradient boosting

A graph network for topology (rings, mule accounts), gradient boosting for account behaviour and anomaly detection for patterns with no labels yet. Their errors are weakly correlated, which is where the gain comes from. Scores are calibrated before combining, and the threshold is set by how many alerts the analyst team can actually review each day.

Clinical RAG with hybrid search | Genetic diagnostics lab | biomedical embeddings

Variant classification against standard clinical criteria over a cohort of more than 60,000 variants. Lexical and vector search run side by side and are fused with a tunable weight, since exact gene names are rare tokens that vector search handles poorly. Chunking keeps tables and section hierarchy intact.

EXPERIENCE

Founder and Lead AI Engineer | GLEMES FFT Sep 2025 to Present | Madrid

- Independent practice with full client ownership: scoping, architecture, build, deployment and follow-up for clients in pharma, genetic diagnostics, insurance and financial services. Up to ten projects in parallel.
- Agentic document ingestion into a canonical schema. The agent proposes a field mapping, validation runs again, and after a fixed number of attempts the case is routed to a human reviewer.
- Multimodal ingestion of expert panel recordings and slide decks into semantic search. Video is cut into overlapping segments so no sentence is lost at a boundary. A per-session pipeline runs alongside a scheduled global job that picks up themes spanning several sessions.
- Own products: conversational analytics over SQL with a LangGraph supervisor and a read-only query validator; accounting document parsing across six formats including scans, with over 95% classification accuracy, 2.5 s p95 latency and 50 requests per minute.
- Spec-first development workflow on Claude Code, packaged as an npm tool, with git worktree isolation to build several features in parallel.

Senior Data Scientist | ThetaRay Nov 2023 to Sep 2025 | Madrid

- Financial crime detection for banks under regulatory supervision, where every alert had to be explainable to a supervisor and every model had to pass validation.
- Built the alert scoring engine described above, with thresholds tied to the review capacity of the operations team.
- Feature engineering, distribution analysis and drift monitoring in PySpark, orchestrated with MLflow and Airflow.

- Entity resolution at scale using MinHash, LSH and distance metrics.
- LLM tooling that helped analysts get through case files faster, my first production work with LLMs.

Data Scientist | Banco Sabadell Mar 2023 to Nov 2023 | Madrid

- Credit risk and fraud models for retail customers, in a team that had to justify each individual decision to a customer or a regulator.
- SHAP and surrogate models so that a rejection could be explained in plain language.
- Leakage checks on fields only known after a case closes, and model documentation to the standard a regulated bank requires.

Data Scientist | Banco Santander Oct 2020 to Mar 2023 | Madrid

- Risk department. Distributed machine learning over millions of accounts, from raw data to a served endpoint.
- End-to-end pipelines: distributed training on Spark, experiment tracking with MLflow, containerised serving and CI/CD for retraining.
- Geolocation analysis with GeoPandas and GeoSpark for branch coverage, plus automated reports and dashboards that replaced a manual monthly process.

TECHNICAL SKILLS

LLMs and agents: LangGraph, LangChain, Strands, AWS Bedrock AgentCore, MCP (FastMCP), structured outputs, tool calling, human-in-the-loop interrupts, persisted state, step budgets, Vapi (voice agents), LoRA/QLoRA, Hugging Face

Evaluation: Golden datasets, LLM-as-judge with calibration, groundedness and citation recall, abstention rate, regression checks, LangSmith tracing

Backend: Python, FastAPI, SQLAlchemy, REST, async programming, PostgreSQL, SQL, Pytest, multi-tenant SaaS architecture

Automation and infrastructure: n8n, Airflow, Docker, Kubernetes, Terraform, GitHub Actions, GitLab CI, AWS, GCP, Azure, tokenization of sensitive data

Retrieval: Hybrid lexical and vector search, reranking, domain embeddings, pgvector, OpenSearch, ChromaDB, table-aware chunking

Data and ML: PySpark, BigQuery, AWS Glue, MLflow, PyTorch, scikit-learn, XGBoost, graph neural networks, anomaly detection, time series, SHAP, LIME

AI-assisted development: Claude Code in daily use, with git worktrees for parallel work

EDUCATION AND LANGUAGES

M.Sc. Data Science | EOI, Escuela de Organización Industrial | 2020 to 2021

B.Sc. Mathematical Engineering | Universidad Complutense de Madrid | 2015 to 2020

Languages: Spanish (native), English (C1, professional)

Other: Semi-professional hockey player. Five years of clarinet at the conservatory.