

Jose Arbeláez

✉ jose.ancizar.667@gmail.com 📞 +34 722 81 50 03 📍 Madrid

🐙 GitHub 🔗 LinkedIn 🌐 Portfolio

Summary

Data Science & Engineering student focused on LLM systems, agent orchestration and production-grade observability. I build multi-agent applications with LangGraph, LangChain, MCP-managed tools, AgentCards/A2A-style handoffs, RAG over financial and SEC data, streaming execution, and deterministic validation. Currently developing Noetik Recon, a multi-agent M&A research system that turns SEC/EDGAR filings, XBRL, PDFs/HTML and financial APIs into auditable Markdown/HTML reports.

Experience

Data Engineering Intern

Jan 2026 – Present

Naudit HPCN, Madrid

- Operate observability/data pipelines for a banking platform, processing 50+ GB/day with Logstash, Filebeat, Elastic Agents, Elasticsearch, SQL, Grafana and Prometheus
- Built and deployed an end-to-end alarm intelligence system used across several Naudit client projects, adding prioritization to hundreds of daily infrastructure alerts through an Elasticsearch/SQL backend and React/shadcn dashboard
- Developed an anomaly scoring engine with Isolation Forest over 24h and 7-day windows, using out-of-hours activation, historical periodicity deviation and relative frequency to flag high-priority alarms
- Built a pairwise alarm relationship model using conditional co-occurrence probabilities, configurable time windows and SciPy statistical tests; added SHAP interpretability and HistGradientBoosting supervised labeling

President & Co-Founder

Jul 2025 – Present

Google Developer Group on Campus UAM

- Co-founded the chapter and lead AI/cloud events for 100+ students, coordinating a multidisciplinary student team and technical community partnerships
- Organized LLM-focused sessions and a community interview/talk with Diverger leadership; built the official website gdguam.es

Full-Stack Developer

Jun 2024 – Jul 2025

VULCASAN SL, Madrid

- Built a full-stack data management system replacing manual client and inventory workflows across an auto shop network
- Designed backend APIs, session management and CI/CD pipelines in a four-person team

Selected Projects

Noetik Recon / NoetikLab – Multi-Agent M&A Research System

2025 – Present

Python, FastAPI, LangGraph, LangChain, MCP, AgentCards/A2A, RAG

- Architecting an autonomous multi-agent system with orchestrator, extraction, validation and analysis specialists; implemented custom AgentCard/A2A-style discovery, delegation and runtime agent creation
- Integrated MCP-managed tools and providers for SEC EDGAR, XBRL, 10-K/10-Q/S-4/8-K filings, PDFs/HTML, market data and financial APIs; uses a RAG knowledge plane with Pinecone/pgvector-compatible vector backends
- Produces structured Markdown/HTML reports with cited artifacts, verification, self-evaluation and replan loops; includes FastAPI/Next.js cockpit, SSE/WebSocket-style streaming, async event bus, RabbitMQ publisher prototype, OpenTelemetry traces and tests

BSc Thesis: Surrogate Models via Gaussian Processes

2025 – 2026

Python, Bayesian Optimization

- Researching Gaussian Process surrogate models for expensive black-box optimization, including Bayesian inference and acquisition/infill criteria
- Training and evaluating surrogate models on a real-world optimization case

Education

BSc in Data Science and Engineering

2022 – 2026

Universidad Autónoma de Madrid (UAM)

- Focus: machine learning, statistical modeling, generative AI, diffusion SDEs, NLP and reinforcement learning
- Board Member of the University Cybersecurity Committee (1 year)

Google Cloud Engineering Certification

2025 – Present

Google Cloud (Online)

Honors & Awards

- 1st Place – **UAM Investor’s League** among 240+ participants: top Sharpe ratio portfolio (2025)
- Excellence Scholarship – **Community of Madrid** (2024,2025)
- 2nd Place – **Cloud Agent Systems Hackathon**, Google Cloud & Diverger (2025)
- 3rd Place – **UAM Entrepreneurship Hackathon**: AI learning platform for students (2022)

Technical Skills

Languages:	Python, SQL, JavaScript/TypeScript, R, C++, Bash
LLM/Agents:	LangGraph, LangChain, smolagents, MCP, AgentCards/A2A, RAG, prompt engineering, output validation, OpenAI, Gemini, DeepSeek
AI/ML:	Scikit-learn, SciPy, PyTorch, TensorFlow/Keras, SHAP, Bayesian Optimization, Gaussian Processes
Data/RAG:	Elasticsearch, PostgreSQL, pgvector, Pinecone, SQLite, DuckDB, MongoDB, SEC EDGAR/XBRL, ETL pipelines
Observability:	Grafana, Prometheus, Elastic Stack, Kibana, OpenTelemetry, alerting, log/metric pipelines
Infra/Web:	FastAPI, Next.js, React, shadcn/ui, WebSockets/SSE, RabbitMQ, Redis, Docker/Docker Compose, Linux, Git/GitHub, CI/CD, Google Cloud, Vertex AI

Languages

Spanish (Native) English (C1 – Full Professional) French (Basic) Naudit recommendation letter available on request