

Alan Garcia Zambrano

AI Engineer & Data Scientist

alanzambie@gmail.com | Mexico City, Mexico

Open to Remote | Open to Relocation

[linkedin.com/in/alanzambie](https://www.linkedin.com/in/alanzambie) | gitlab.com/alanzambie | zambrians.com

PROFESSIONAL SUMMARY

AI Engineer and Data Scientist, seven years getting machine learning and generative AI all the way into production, which is a much less glamorous job than getting it to work. I have shipped **RAG** pipelines over legal archives nobody was allowed to take home, agentic bots that decide on their own which client portal to visit next, and forecasting models that quietly run somebody's Monday. I usually end up owning the software around the model too, from the **PostgreSQL** schema to the **React** interface, because somebody has to. Recent work spans automotive manufacturing, federal public health at national scale, and electoral institutions, on **Azure**, **AWS** and **Databricks**.

PROFESSIONAL EXPERIENCE

- 02.2026 - Present **AI Engineer & Data Scientist (Consultant)**
Act Digital, Mexico City, Mexico (Remote)
- **Demand Forecasting:** Improved demand forecasting for automotive glass components at GM, Ford and Stellantis, so planning cycles got shorter and inventories got tighter. Nothing exotic underneath, just time series and regression on data I extract and clean myself, which is the half nobody puts in the slide deck.
 - **Agentic AI, Automation & Monitoring:** Client data lived behind a dozen portals, each with its own personality, so I built autonomous bots that work out on their own how to get through them, scheduled as **Databricks** jobs. A **React** dashboard on **Docker** monitors the health of the whole fleet, with every run, its status and the daily reports in one place. A workflow that used to eat days now runs from one trigger, and the team only looks when the dashboard says something broke.
 - **Data Architecture & MLOps:** I own the model lifecycle across **Databricks** and **Azure**, with distributed processing on **Python** and **PySpark** and versioning and **CI/CD** on GitLab, from data prep through training and evaluation. It is the least glamorous part of the job and it is why the models are still right six months later.
 - **Stakeholders & Reporting:** I sit with the business owners and with IT to turn tangled demand into forecasts they will actually use. Delivery is where most models quietly die, so these ones send themselves as daily executive reports, emails, PDFs and Office files. Nobody there has had to ask for a report since, which I count as the real win.
- 01.2024 - 01.2026 **Data Scientist / Head of Impact Analysis (Consultant)**
Ministry of Health, Public Health Policy, Mexico City, Mexico
- **National Platform, High Concurrency:** Led the Semana Nacional de Salud Pública platform from first commit to national rollout, twice, in 2025 and again in 2026. The **React** and **PostgreSQL** stack, containerized with **Docker** and monitored with **Grafana**, held up under 3,000+ concurrent users through the peak weeks and collected structured and unstructured records from more than 20 million citizens. I never had a harder week than that one, watching more numbers and charts on that dashboard than I care to remember.
 - **Automated Web Ecosystem:** A second platform (**React**, **PostgreSQL** on **Docker**, Google Drive API) controls national health records, with hierarchical permissions for 200+ federal and state users. From there nothing depends on a person, because incoming data feeds the ARIMA forecasts, backups and reports go out on schedule, and medical inventories get certified nationwide. Building it so nobody had to step in was the assignment.
 - **Containers, Bots & BI:** Modernized the internal health certification platform and moved production **PostgreSQL** onto containers in **OpenShift**. Manual data entry disappeared the week I finished the Playwright and **Python** bots that fill fields and register users in bulk, and out the other end come the daily health indicators in the **Tableau** and **React** dashboards the executive team uses.
- 01.2025 - 12.2025 **Artificial Intelligence Specialist (Consultant)**
National Electoral Institute, Mexico City, Mexico
End-to-end **RAG** for legal and budgetary analysis
- **Production RAG Architecture:** Analysts had been searching these archives by hand, one folder at a time, so I shipped a production **RAG** system over highly sensitive legal and budgetary documents that pulls exact figures and clauses instead. The documents could not leave the building, so it runs on self-hosted open models, **DeepSeek** and **Qwen**, with **Python** and **LangChain**.
 - **Hybrid OCR Pipeline:** Accuracy depended entirely on the ingest pipeline, so I built a hybrid one around Google Document AI that recovers text and spatial layout from scanned legal files, tables and multi-column pages included. Scanned government paperwork is its own kind of adversary.
 - **Retrieval & Vector Storage:** Tuned retrieval until hallucinations stopped being a problem, evaluating chunking strategies (hybrid, semantic, sliding window) against real queries while embeddings and high-volume vector search ran on SurrealDB. On legal documents a confident wrong answer is worse than no answer.
 - **Full-Stack Productization:** Built the full-stack app on top (**React**, **Python**) so non-technical staff can query the archive in plain language instead of going through IT. Watching a lawyer get an answer assembled from more than five documents, that fast, made the project worth it.

01.2022 - 12.2024	Data Scientist GO-SHARP, Mexico City, Mexico (Remote) ▪ Technical Leadership & ML: Led the machine learning team on retail accounts like Nestlé, Walmart, Sigma and chains of pharmacies and gas stations, with catalogues of hundreds of thousands of SKUs spread over hundreds of stores, which is where forecasting stops being statistics and turns into engineering. I covered both supervised forecasting and unsupervised pattern discovery in Python, PySpark, scikit-learn, TensorFlow and PyCaret, with XGBoost, Random Forest and K-Means. The hard part was never the model, or even cleaning the data, it was convincing the client. ▪ Cloud Data Platform: Set up the data platform across Azure, Databricks and AWS Redshift, with ETL running on Airflow over Bash and cron. Once it was tuned the whole high-volume ingestion path ran unattended, which is about the highest compliment infrastructure ever gets. ▪ Statistics & BI: The statistics were mine too, from RFM and market basket analysis to hypothesis testing and bootstrapping, in Power BI and Tableau dashboards the sales team opened every morning.
01.2020 - 01.2023	Data Scientist (Consultant) Epidemiological and Health Intelligence Unit, Mexico City, Mexico (Remote) ▪ Modeling, NLP & Statistical Testing: Wrote the models behind national containment decisions, both time series and regression forecasts and an NLP system that read social media sentiment to flag public health risks. It all ran in Python and R, and every result went through ANOVA and t-tests first, because the cost of being wrong was not a bad quarter. ▪ Reporting Automation: Before that, the daily report was assembled by hand every morning. I automated the whole cycle, coordinating SQL extraction, transformation and the operational reports with Python and Bash, while geospatial dashboards in Tableau and Leaflet showed how the disease was spreading.
01.2019 - 12.2019	Big Data Specialist Total Play, Mexico City, Mexico ▪ Big Data Engineering & ETL: Built and cleaned the ETL pipelines feeding the analytics team, in Python, SQL and Boto3 over Linux clusters, AWS S3 and Redshift. That job taught me how much of this work is janitorial. ▪ Data Mining & BI: Found unclaimed revenue in subscriber behaviour using market basket analysis and segmentation, and published it in automated Tableau, QlikSense and Power BI dashboards. This was 2019, before ChatGPT existed to explain a complex report for you.

EDUCATION

2020 - 2022	Master of Business Administration (MBA) Universidad de las Tres Culturas (UTC), Mexico City
2017 - 2019	M.Sc. in Computer Technology (Artificial Intelligence) Instituto Politécnico Nacional (IPN), Mexico City
2009 - 2016	B.Sc. in Computer Engineering Instituto Politécnico Nacional (IPN), Mexico City

ACHIEVEMENTS & RECOGNITION

2025	National Digital Health Infrastructure Scaled the platform behind Mexico's Semana Nacional de Salud Pública, which held up through two consecutive editions and served millions of citizens nationwide.
2021	2nd Place, Datatón México Awarded for a high-impact data solution in a competitive national hackathon.
2019	Generative Music Composer Machine learning algorithm that composes original music, built years before generative AI went mainstream.
2019	Academic Participant, McGill University Quebec, Canada.

TECHNICAL SKILLS

- **Programming Languages** - Python, R, SQL, Bash, JavaScript, TypeScript, C++
- **AI, GenAI & Machine Learning** - RAG, LLMs, Agentic AI, LangChain, Embeddings, Vector Search, RAG Evaluation, OpenAI, Claude, Gemini, DeepSeek, Qwen, NLP, OCR (Google Document AI), Vector Databases (SurrealDB), TensorFlow, Scikit-learn, PyCaret, XGBoost, Random Forest, Time Series (ARIMA)
- **Data Engineering & Big Data** - Databricks, PySpark, Airflow, ETL, PostgreSQL, AWS Redshift, Web Scraping (Playwright), Boto3, Pandas, NumPy
- **Cloud & MLOps** - Azure, AWS (EC2, S3), MLOps, Docker, OpenShift, CI/CD, Git (GitHub, GitLab), Grafana, Linux
- **Visualization & Full-Stack** - Tableau, Power BI, React, Next.js, Shiny (R), ggplot2, Leaflet
- **Analytics & Methodologies** - Hypothesis Testing, ANOVA, Bootstrapping, RFM, Market Basket Analysis, Agile/Scrum, JIRA

LANGUAGES

Spanish - Native | English - Professional working proficiency (C1)