

# Carlos Alberto Haro López

## AI Engineer & Technical Product Manager

carlos@h1sort.com · h1sort.com · github.com/haro-ca · linkedin.com/in/h1sort · x.com/h1sort · Mexico City, MX

8+ years of experience designing and implementing end-to-end data & AI solutions across AWS, Azure, and GCP.

**Adept at bridging technical execution and business strategy to drive innovative and scalable data solutions, for sales and for execution.**

---

### HIGHLIGHTS

- **Current role** — Santander AI: AI Engineer & Technical Product Manager, building and shipping AI systems for one of Mexico's largest banks, end to end.
- **Ships agents to production** — FRED, a structured fraud-analysis agent live inside Santander Mexico: 10k credit petitions analyzed in 1 hour, replacing a ~1k-per-week manual bottleneck.
- **Diverse background as an individual contributor** — Data Engineering (2 years), Data Science (2 years), ML Engineering (1 year), and Software Engineering (1 year), providing a comprehensive perspective across the full data lifecycle.

---

### RECENT EXPERIENCE

#### Santander AI — AI Engineer & Technical Product Manager 2025 - PRESENT · MEXICO CITY

##### RAG Chatbot · *Technical PM*

- **Problem:** High cost-to-serve for clients caused by scattered knowledge around retail financial product characteristics and procedures.
- **Product:** Cloud RAG system for the branch network & contact center — Knowledge Base Owners management, prompt design, and evals strategy design and implementation.
- **Highlights:** Evals scaled from <100 to >1,000 while cutting eval time 10x through improved strategy.
- **Outcome:** V1.0 in production at 20% adoption — V2 shipping next.

##### FRED — structured fraud-analysis agent · *AI Engineer*

- **Problem:** Manual fraud analysis capped at ~1k credit petitions per analyst per week.
- **Product:** Deterministic fraud-analysis pipeline over financial signals. Full ownership — architecture through deployment, including the FRED name & identity.
- **Outcome:** In production at Santander Mexico — 10k credit petitions analyzed in a single hour.

##### ProntoGPT — internal agent harness · *AI Engineer*

- **Problem:** A budget constraint left only 10% of the Mexico organization with a ChatGPT license.
- **Product:** Prototype Santander agent harness demonstrating org-wide AI access.
- **Outcome:** Influenced the global organization to make AI available to every Santander Mexico employee — today delivered through Copilot Chat.

##### AI teaching & Knowledge Base MVP · *AI Engineer*

- Taught 300+ employees across 6 bank areas on agent harnesses (ChatGPT, Copilot, Claude, Windsurf, Devin, Cortex); built internal AI tooling now used by 50+ engineers.
- Knowledge-base agent with two tools — structured SQL + unstructured OCR RAG — being deployed to production AWS as KB v0.1.

## Cloud MLOps Framework — international bank · *PM, Solution Architect, Lead Engineer* · 2024–25

- **Problem:** Data science area capable of model development and deployment but facing slow iteration speed due to inexistent CI/CD (MLOps) pipelines.
- **Product:** PaaS cloud MLOps configuration and deployment (Snowflake ML, VertexAI, Sagemaker, AzureML, Databricks); migration of models from legacy proprietary autoML to open-source equivalents, matching model results and architecture.
- **Outcome:** V1.0 delivered to production.

## Analytics Platform Assessment — international insurer · *Lead Engineer* · 2023

- **Problem:** Data platform modernization efforts were giving low business results — high SLAs for incorporating new data due to poor relational modelling and a missing analytics-engineering framework.
- **Product:** Prioritized optimization roadmap from a full data value-chain assessment — OnPrem transactional DBs, DWH, ongoing cloud migration, and all downstream BI consumption.
- **Outcome:** Delivered; adopted into implementation plans.

## Omnichannel Unified Data Model — international retailer · *Lead Engineer, Solution Architect* · 2022

- **Problem:** Brick & mortar and e-commerce KPIs were consulted through different methodologies and software, joined through ad-hoc procedures giving different results per area.
- **Product:** Unified data model on the central cloud data warehouse, exposed via BI dashboard; ETLs designed and implemented from scratch in PySpark.
- **Outcome:** Delivered — ETL pipelines, unified relational model, and KPI semantic layer still in production use.

## Prior roles

Software Engineer — Development Bank (2020–21) · Data Scientist — Central Tax Administration (2018–20) · Data Engineer — Boutique consulting (2017)

---

### TEACHING

**Real-Time Data Processing** — a 12-lesson hands-on course, from a single Postgres node to a full streaming pipeline (OLTP → CDC → Kafka → stream processing → real-time OLAP → API). [github.com/haro-ca/real-time-data-processing-class](https://github.com/haro-ca/real-time-data-processing-class)

---

### TECHNICAL STACK

- **AI & Agents:** RAG systems · vector search · agent frameworks · evals · prompt engineering · MCP · LLM deployment & monitoring
- **MLOps & Platforms:** VertexAI · Sagemaker · AzureML · MLflow · Databricks · Snowflake ML · CI/CD for ML
- **Data Engineering:** Databricks · Snowflake · BigQuery · Azure Synapse · Redshift · DataFactory · EMR/Glue · DataFlow · Airflow · dbt · dagster
- **Cloud:** AWS · Azure · GCP — 8+ years multi-cloud
- **BI & Analytics:** Looker · Power BI · semantic layer design
- **Fullstack (limited):** Next.js · Astro · FastAPI · Flask · Python · PySpark · SQL