

Dear Hiring Manager,

I am an AI/NLP Solutions Architect with a linguistics background who specializes in turning messy business questions into retrieval-first, auditable AI systems. In my recent work and at Spotify, I have designed and shipped end-to-end pipelines that connect natural-language interfaces to reliable data and decisions, with safety, latency, and cost guardrails built in from day one.

Most recently, I designed an end-to-end architecture for a Retail Copilot that lets business users ask natural-language questions and receive validated SQL and charts over BigQuery. I authored a 35+ page architecture dossier that lays out a path from PoC to MVP to multi-tenant deployment on GCP/Vertex, using API Gateway, Cloud Run, Vertex AI, and BigQuery. The core of the system is an NL→intent→slots→template→validator chain with allowlists, tenant filters, dry-run bytes, and budget caps—so every query is explainable, bounded, and tied to explicit contracts rather than ad-hoc prompts.

Because this copilot is intended to serve multiple retail tenants, I designed tenancy and security as first-class concerns: per-tenant datasets, logs, budgets, and service accounts, with a clear blast radius and impersonation model. Around the copilot, I specified an MLOps and evaluation harness: golden sets, promotion gates, canary rollout and rollback policies, and monitoring hooks for quality, drift, and cost. I am happy to walk through this dossier or the corresponding GitHub repo in a call.

At Spotify, I built a privacy-compliant NLP classification pipeline for GDPR/CCPA workflows, translating regulatory constraints into taxonomy, data, and evaluation requirements. I also delivered a semantic retrieval capability using transformer embeddings and FAISS for internal knowledge search, defining indexing, freshness, and safety policies. Across these projects, my role has been to partner with Legal, Operations, and Product to turn policy and business intent into system guarantees—labeling strategies, redaction paths, safety thresholds, and regression tests that can be audited.

What I bring to an AI Solutions Architect role is a problem-to-production approach:

- Frame ambiguous business goals as concrete retrieval, classification, or NL→SQL problems with measurable success metrics.
- Map those problems to proven patterns: hybrid RAG, semantic search, tool-using LLMs, and agentic review loops rather than one-off prompts.
- Design and implement the pipeline: APIs and services (FastAPI/Cloud Run), data and retrieval layers (BigQuery, FAISS/HNSW), and orchestration/CI (Docker, GitHub Actions, MLflow/Flyte).
- Wrap the system with evaluation and safety: golden sets, offline/online metrics, cost/latency budgets, and monitoring that lets stakeholders trust what the system is

doing.

I am particularly interested in roles where I can own the architecture of NL-driven workflows—RAG systems, NL→SQL copilots, agentic evaluators—and work closely with both technical and non-technical stakeholders to ensure the solution is understandable, maintainable, and aligned with business impact. My GitHub portfolio (github.com/naaas94) showcases these patterns in working code, from privacy-compliant classifiers to agentic review systems and data pipelines.

I would welcome the opportunity to discuss how this mix of architecture, hands-on implementation, and stakeholder partnership could support your AI initiatives.

Best regards,
Alejandro Garay

Buenos Aires, Argentina
alejandroa.garay.ag@gmail.com · +54 11 5660 5331
GitHub: github.com/naaas94 · LinkedIn: linkedin.com/in/alejandro-garay-frontini