

Ariel Bailon Intriago

AI Engineer · Full-Stack Engineer (backend) · Agents, RAG and evals on Node.js/NestJS and Python

+593 985 267 399 | arielbailonn@outlook.es | Manta, Ecuador (Remote, UTC-5) | [Portfolio](#) | [LinkedIn](#) | [GitHub](#)

CAREER SUMMARY

Backend-focused Full-Stack Engineer shipping production Node.js/NestJS backends since 2020, now building LLM systems end to end: tool-calling agents, cited RAG and automated evals. My latest project, OrderOps Agent, is a NestJS + LangGraph.js support agent where the model proposes and the code decides: refund rules are enforced in code, every refund goes through human approval, and quality is measured by 41 eval cases in CI. Production track record of cutting API response times (~35%) and load times (~25%). Looking for an AI Engineer or AI-focused backend role where I can take features from problem to production and measure whether they work.

TECHNICAL SKILLS

Agents & LLMs: LangGraph.js, LangChain.js, tool/function calling, human-in-the-loop, multi-turn memory, Anthropic Claude, OpenAI, AWS Bedrock, OpenRouter, provider failover, prompt caching, structured outputs (Zod, Pydantic), SSE streaming

RAG: embeddings (sentence-transformers), vector search (pgvector, ChromaDB), section-aware chunking, cited answers, refusal guardrails

Evaluation & quality: LLM-as-judge evals, adversarial cases, regression checks in CI (GitHub Actions), cost/token/latency tracking, Vitest, pytest

Backend: Node.js, NestJS, TypeScript, Python, FastAPI, PostgreSQL, MongoDB, MySQL, GraphQL, REST, Prisma, Redis, BullMQ

Frontend & infrastructure: React, Next.js, Angular · Docker, AWS (EC2, S3, Bedrock), NGINX, DigitalOcean, Git

AI-assisted delivery: Claude Code with a spec-first workflow (spec → implement → test/type/lint checks → fix records); 31 specs and 15 fix records across 3 public repos

AI PROJECTS — END-TO-END OWNERSHIP

OrderOps Agent | github.com/ArielBailon/order-ops

Oct. 2026

E-commerce support agent that answers policy questions, looks up orders and proposes refunds a human approves — solo build

- Designed a LangGraph.js state graph with an intent router and 4 routes: store policies (RAG over pgvector), order lookups (tool calling with the user's identity, so each customer only sees their own orders), refunds, and controlled out-of-scope refusal.
- Separated generative from deterministic logic: refund rules live in code, not the prompt, and the agent pauses (interrupt) until an operator approves or rejects each refund.
- Put Anthropic, OpenAI and AWS Bedrock behind one interface with provider failover and prompt caching; per-conversation memory (PostgreSQL checkpoint) and per-customer memory.
- Built a 41-case eval suite (13 adversarial) scoring tool selection, guardrail behavior and answer quality with an LLM judge; runs in CI with a cost cap and regression check. 261 automated tests.

Technologies: TypeScript, NestJS, LangGraph.js, LangChain.js, PostgreSQL + pgvector, Prisma, Zod, Vitest, GitHub Actions

EDGAR Analyst | github.com/ArielBailon/edgar-analyst

Sep. 2026

Grounded Q&A over SEC filings and earnings calls, citing the exact filing and section — solo build

- RAG pipeline over 10-K/10-Q filings and transcripts: section-aware chunking (~2,400 cited chunks), semantic search with ChromaDB and Claude Haiku 4.5 answers with exact citations.
- Pre-LLM refusal guardrail: no supporting evidence, no answer, so off-topic questions cost \$0.
- Eval set of 22 hand-labeled pairs (citation accuracy and relevance); diagnosed the gap as a retrieval problem, not a generation one, and documented the improvement hypotheses.
- Per-query logging of latency, tokens and cost (~2 s median, ~\$0.002 per answer); 122 deterministic pytest tests.

Technologies: Python, Anthropic API, ChromaDB, sentence-transformers, Pydantic, pytest

PROFESSIONAL EXPERIENCE

Independent — AI Engineering

Jul. 2026 – Present

Own AI projects: OrderOps Agent and EDGAR Analyst (above) and CodeSage

- **CodeSage (in progress):** assistant that answers questions about a code repository. FastAPI microservice with SSE streaming, Pydantic-validated output, exponential-backoff retries, per-request cost tracking and versioned prompts; 74 tests. Next: RAG with pgvector, hybrid search and file-and-line citations.

Technologies: Python, FastAPI, Pydantic, PostgreSQL + pgvector, TypeScript, NestJS, Docker

ULTAHOST, United States

Aug. 2025 – Oct. 2025

Backend Developer (Contract)

- Designed and implemented scalable backend APIs with NestJS and TypeScript, improving response times by ~35% through query optimization and caching.
- Built and maintained REST and GraphQL services integrated with internal and third-party systems.
- Implemented a microservices architecture for high-availability systems on MySQL.

Technologies: NestJS, TypeScript, MySQL, GraphQL

MUSHROOMSOFT-IT, Quito, Ecuador

Jun. 2023 – Jan. 2024

Full Stack Developer

- Developed backend services for a micro-frontend architecture (NestJS + Angular), cutting load times by ~25% through PostgreSQL indexing and query optimization.
- Designed data models in PostgreSQL and MongoDB for a system with 50k+ active records.
- Containerized applications with Docker, standardizing environments and eliminating configuration incidents across the team.
- Led weekly code reviews, establishing quality standards adopted by a team of 5 developers.

Technologies: NestJS, Angular, PostgreSQL, MongoDB, Docker

ZGAMESLATAM, Manta, Ecuador

Jun. 2022 – Jun. 2023

Full Stack Developer

- Designed and implemented REST APIs in Node.js for an internal ERP processing ~200 daily transactions, improving data consistency across modules.
- Optimized NGINX server configuration, achieving a ~20% reduction in average latency.
- Managed deployments on AWS and DigitalOcean for a team of 4.

Technologies: Node.js, Angular, AWS, DigitalOcean, NGINX

CONSULTI, Guayaquil, Ecuador

Sep. 2020 – Jun. 2022

Junior Full Stack Developer

- Developed backend APIs in Node.js and Spring Boot for an e-commerce platform with 5,000+ products and multiple integrated payment methods.
- Deployed applications on AWS and DigitalOcean using Docker and NGINX, reducing deploy time by ~40%.
- Applied clean code and Scrum in a team of 6, consistently meeting delivery deadlines.

Technologies: Node.js, Spring Boot, Docker, AWS, NGINX

EDUCATION

Universidad Laica "Eloy Alfaro" de Manabí, Manta, Ecuador

2014 – 2021

BS, Systems Engineering

LANGUAGES

Spanish (native — C2) | English (advanced — C1)