

Ariel Bailon Intriago

AI Engineer · Ingeniero Full-Stack (backend) · Agentes, RAG y evals sobre Node.js/NestJS y Python

+593 985 267 399 | arielbailonn@outlook.es | Manta, Ecuador (Remoto) | arielbailon.dev | [LinkedIn](#) | [GitHub](#)

PERFIL PROFESIONAL

Ingeniero Full-Stack con enfoque backend: entrego backends en producción con Node.js/NestJS desde 2020 y hoy construyo sistemas con LLMs de punta a punta: agentes con tool calling, RAG con citas, evaluaciones automáticas y observabilidad. Mi proyecto más reciente, OrderOps Agent, es un agente de soporte en NestJS + LangGraph.js donde el modelo propone y el código decide: las reglas de reembolso se validan en código y cada reembolso pasa por aprobación humana. Está desplegado en AWS con CI/CD, trazas por request y una demo pública (demo.arielbailon.dev), y su calidad se controla con **41 casos de evaluación** como gate en CI. Historial en producción reduciendo tiempos de respuesta de APIs (~**35%**) y de carga (~**25%**). Busco un rol de AI Engineer o backend con IA donde pueda llevar funcionalidades desde el problema hasta producción y medir si funcionan.

HABILIDADES TÉCNICAS

Agentes y LLMs: LangGraph.js, LangChain.js, tool/function calling, human-in-the-loop, memoria multi-turno, Anthropic Claude, OpenAI, AWS Bedrock, OpenRouter, failover entre proveedores, prompt caching, salidas estructuradas (Zod, Pydantic), streaming SSE

RAG: embeddings (sentence-transformers), búsqueda vectorial (pgvector, ChromaDB), chunking por sección, respuestas con citas, guardrails de rechazo

Evaluación y observabilidad: evals con LLM-as-judge, casos adversariales, regresión como gate en CI, trazas con Langfuse, seguimiento de costo, tokens y latencia por request, Vitest, pytest

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

Frontend, infraestructura y CI/CD: React, Next.js, Angular · Docker, AWS (EC2, ECR, SSM, IAM, CloudWatch, S3, Bedrock), Caddy, NGINX, DigitalOcean · GitHub Actions (CI/CD con OIDC)

Desarrollo asistido por IA: Claude Code con flujo spec-first (spec → implementación → checks de tests, tipos y lint → registro de fixes); 31 specs y 15 fix records documentados en 3 repos públicos

PROYECTOS DE IA — OWNERSHIP END-TO-END

OrderOps Agent | demo.arielbailon.dev | github.com/ArielBailon/order-ops

Oct. 2026

Agente de soporte para e-commerce que responde políticas, consulta pedidos y propone reembolsos con aprobación humana — desarrollo individual, en producción con demo pública

- Diseñé un grafo de estados en LangGraph.js con un router de intención y **4 rutas**: políticas de la tienda (RAG sobre pgvector), consulta de pedidos (tool calling con la identidad del usuario: cada cliente solo ve sus propios pedidos), reembolsos y rechazo controlado de lo que está fuera de alcance.
- Separé lo generativo de lo determinista: las reglas de reembolso viven en código y no en el prompt, y el agente se pausa (interrupt) hasta que un operador aprueba o rechaza cada reembolso. Anthropic, OpenAI y AWS Bedrock detrás de una sola interfaz con failover y prompt caching.
- Lo desplegué en AWS: EC2 con Caddy (TLS) y Bedrock vía IAM role, sin access keys en el servidor. CI/CD con GitHub Actions: OIDC, imágenes en ECR etiquetadas por commit, deploy por SSM, smoke test y rollback automático. La demo pública tiene rate limits, tope de gasto diario y reset nocturno.
- Observabilidad: cada request queda trazado en PostgreSQL y Langfuse (ruta, herramientas, proveedor, failover, tokens, costo y latencia) y se muestra en una consola Next.js con timeline del agente y dashboard p50/p95; ~**US\$0.012** por conversación.
- 41 casos de evaluación (13 adversariales)**, **41/41** en la última corrida en producción, usados como gate de regresión en cada pull request; **457 tests automatizados**.

Tecnologías: TypeScript, NestJS, LangGraph.js, LangChain.js, PostgreSQL + pgvector, Prisma, Zod, Next.js, Langfuse, AWS (EC2, ECR, SSM, IAM, CloudWatch, Bedrock), Docker, GitHub Actions

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

Sep. 2026

Preguntas y respuestas fundamentadas sobre reportes de la SEC y earnings calls, con cita al reporte y sección exactos — desarrollo individual

- Pipeline RAG sobre reportes 10-K/10-Q y transcripciones: chunking por sección (~**2,400 chunks citables**), búsqueda semántica con ChromaDB y respuestas de Claude Haiku 4.5 con cita exacta.
- Guardrail de rechazo previo al LLM: sin evidencia suficiente no hay respuesta, y las preguntas fuera de tema cuestan **\$0**.
- Evaluación con **22 pares etiquetados a mano** (precisión de citas y relevancia); diagnostiqué que la brecha está en el retrieval y no en la generación, y documenté las hipótesis de mejora.

- Logging por consulta de latencia, tokens y costo (mediana ~2 s, ~\$0.002 por respuesta); **122 tests deterministas** con pytest.

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

EXPERIENCIA PROFESIONAL

Independiente — AI Engineering

Jul. 2026 – Actualidad

Proyectos propios de IA: OrderOps Agent y EDGAR Analyst (detalle arriba) y CodeSage

- **CodeSage (en progreso)**: asistente que responde preguntas sobre un repositorio de código. Microservicio FastAPI con streaming SSE, salida validada con Pydantic, reintentos con backoff exponencial, costo por request y prompts versionados; **74 tests**. En curso: RAG con pgvector, búsqueda híbrida y citas por archivo y línea.

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

ULTAHOST, Estados Unidos

Ago. 2025 – Oct. 2025

Backend Developer (Contrato)

- Diseñé e implementé APIs backend escalables con NestJS y TypeScript, mejorando los tiempos de respuesta en ~35% mediante optimización de consultas y caché.
- Construí y mantuve servicios REST y GraphQL integrados con sistemas internos y de terceros.
- Implementé una arquitectura de microservicios para sistemas de alta disponibilidad con MySQL.

Tecnologías: NestJS, TypeScript, MySQL, GraphQL

MUSHROOMSOFT-IT, Quito, Ecuador

Jun. 2023 – Ene. 2024

Full Stack Developer

- Desarrollé servicios backend para una arquitectura de micro-frontends (NestJS + Angular), reduciendo tiempos de carga en ~25% con índices y optimización de consultas en PostgreSQL.
- Diseñé modelos de datos en PostgreSQL y MongoDB para un sistema con más de **50k** registros activos.
- Contenericé aplicaciones con Docker, estandarizando entornos y eliminando incidentes de configuración en el equipo.
- Lideré code reviews semanales, estableciendo estándares de calidad adoptados por un equipo de 5 desarrolladores.

Tecnologías: NestJS, Angular, PostgreSQL, MongoDB, Docker

ZGAMESLATAM, Manta, Ecuador

Jun. 2022 – Jun. 2023

Full Stack Developer

- Diseñé e implementé APIs REST en Node.js para un ERP interno con ~200 transacciones diarias, mejorando la consistencia de datos entre módulos.
- Optimicé la configuración de NGINX, logrando ~20% de reducción en la latencia promedio.
- Gestioné despliegues en AWS y DigitalOcean para un equipo de 4.

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

CONSULTI, Guayaquil, Ecuador

Sep. 2020 – Jun. 2022

Junior Full Stack Developer

- Desarrollé APIs backend en Node.js y Spring Boot para un e-commerce con más de **5,000 productos** y múltiples métodos de pago integrados.
- Desplegué aplicaciones en AWS y DigitalOcean con Docker y NGINX, reduciendo el tiempo de despliegue en ~40%.
- Apliqué clean code y Scrum en un equipo de 6, cumpliendo consistentemente los plazos de entrega.

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

EDUCACIÓN

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

2014 – 2021

Ingeniería en Sistemas

IDIOMAS

Español (nativo — C2) | Inglés (avanzado — C1)