

Alexis Aguirre

AI Engineer / Full-Stack Developer - LLM products, RAG, autonomous agents

Cordoba, Argentina (Remote) | aguirrealexis.cba@gmail.com | linkedin.com/in/alexisaguirre-aideveloper | github.com/ale-aguirre

Open to opportunities, Full-time or Contract

SUMMARY

AI-focused Full-Stack Developer with 5+ years shipping production apps and 2+ years building LLM-powered products end-to-end. Designs prompt pipelines, RAG over real domain corpora, and autonomous browser agents. Owns the whole stack: data ingestion, vector stores, model orchestration, evaluation, frontend, and deployment. Comfortable picking the right model (Claude, Groq, OpenAI) for cost and latency, not by hype.

AI / LLM

Models: Claude (Opus, Sonnet, Haiku), Groq Llama 3.3 70B, OpenAI GPT-4/5, Anthropic SDK, OpenAI-compatible proxies

Prompting: system prompt design, structured output (JSON), function calling, tool use, prompt caching, eval harnesses

RAG: ChromaDB, LanceDB, pgvector, hybrid search, chunking, doc parsing (PDF/DOCX), citation tracking

Agents: browser agents (Playwright+CDP), MCP servers, multi-agent orchestration, Claude Code / Agent SDK, vision (Claude/OpenAI), OCR pipelines

FULL-STACK

Frontend: React, Next.js (Pages and App Router), TypeScript, Tailwind, shadcn/ui, Framer Motion, Three.js / R3F

Backend: Node.js, Python, FastAPI, Flask, REST + streaming APIs, Supabase, PostgreSQL, Prisma, SQLite

Tooling: Git, Claude Code, Cursor, Playwright, Puppeteer, Vercel, Docker, Railway

EXPERIENCE

Full-Stack Developer

2021 - Present

Contenidos Digitales, Digital agency serving LATAM and international clients across real estate, hospitality, and brand projects. Outsourced team on campaign work inside Spotify's ecosystem, including brand activations with Mastercard and Visa.

- Built and shipped 10+ production React/Next.js sites for real estate, hospitality, and brand clients, with auth, admin panels, and Google Maps integrations.
- Built responsive multi-language interfaces with complex form validation (Formik/Yup), Framer Motion animations, and CMS-driven content via Strapi.
- Integrated GraphQL (Apollo Client) across all active client projects - nested queries, mutations, fragments, and server-side rendering in Next.js.
- Shipped multi-language production sites on tight agency timelines, coordinating with design and backend across concurrent client projects.

AI PROJECTS

DocUnify - Document Reconciliation Tool (side project for a client, on hold)

Next.js, Python, Flask, Groq Llama 3.3 70B, ChromaDB, Railway.

- Python/Flask backend that ingests, extracts, matches and compares Word documents, then generates a unified DOCX with the LLM recommendation applied.
- RAG pipeline over a standards corpus with ChromaDB vector store, tunable match threshold, and Groq API key rotation on rate limits.

Forgix - AI-Powered Multiplayer Robot Battle Game

Next.js 14, React Three Fiber, Rapier, Supabase, Claude API, Groq SDK. URL: forgix.xyz

- LLM-driven NPC personalities with persistent memory across battles, async PvP with server-side simulation, ELO matchmaking, in-game P2P marketplace.
- Operator League slice: hex-stat cognitive chip system that maps to prompt parameters for AI companions. 24,000+ lines of TypeScript, deployed.

job-hunter - Autonomous Job Application Agent (open source)

Node.js, Playwright, Claude API, Groq, SQLite. github.com/ale-aguirre/claude-job-hunter

- Multi-agent pipeline that scouts 17 job board APIs, scores fit, tailors a CV per posting against a fixed fact registry so the model cannot invent experience, and submits through the ATS.
- 316 applications submitted autonomously, 248 verified on the target page. Grading delivery per channel instead of in aggregate exposed one integration failing silently 100% of the time while the overall number looked healthy.
- Promptfoo eval suite over the CV pipeline. It caught a prompt injection embedded in a job description that made the model claim ten years of Kubernetes; the deterministic validator stopped it from reaching the document.
- A QA auditor pauses a worker after three failures and resumes it on a clean cycle, so one broken integration does not take down the run.

intent-gate - Deterministic Routing for LLM Agents (open source)

TypeScript, Vitest, zero dependencies. github.com/ale-aguirre/intent-gate

- The model classifies a message into one of a known set of domains; which executor runs and whether a human must approve come from configuration the model never sees.
- Text arriving in the input cannot name a handler that does not exist, nor argue its way past an approval gate. Silence on an approval resolves to denied, never to allowed. 25 tests, CI on Node 20 and 22.

EDUCATION & LANGUAGES

Henry, Full-Stack Development Program (2020 - 2021). Spanish: Native. English: Upper-intermediate (professional writing, conversational).