

Anuar Harb

Forward Deployed AI Engineer · LLM Agents in Production

Barranquilla, Colombia · open to relocation Mexican & Colombian citizen Spanish & English

anuarharb@gmail.com +57 310 390 0986 anuarharb.com linkedin.com/in/anuarharb github.com/AnuarHarb

AI engineer with ten years shipping production software and three years taking LLM agents from a company's real problem to production. I work embedded with the people I build for: discovery with owners and operators, pilots with success metrics agreed upfront, integration into their own systems, and measurement after launch. I'm as comfortable with a customer's engineers as with their executives, and I build on one rule: the model talks; the code decides anything with consequences.

SELECTED WORK

- **Sofia, collections voice agent for a demo lender.** Voice agents for two lending brands built from one base template and validated against the ElevenLabs OpenAPI schema before every deploy; both pass the same 22-test battery. A FastAPI integration layer enforces identity, amounts and Colombian collections law (Ley 2300) in code, with Postgres row-level security per brand and a console with every call's audit trail. github.com/AnuarHarb/voice-agent
- **Agents in client production.** A legal intake assistant with 300+ recurring users, a quoting agent for an interior design studio saving ~2 hours per quote, a real estate assistant that answers from live inventory and books viewings, and a student enrollment agent, all over WhatsApp Cloud API and wired into each client's systems. Discovery to production in 2 to 3 weeks.
- **Quillami Code, an open-source coding agent.** ~1,400 lines of TypeScript on the raw Anthropic API, no orchestration framework: agent loop, tools, permission gating, checkpoints, context compaction, a programmatic eval suite and CI. Published on npm.

EXPERIENCE

Forward Deployed AI Engineer & Technical Lead · Inmercity Labs (applied AI studio I co-founded) Jan 2026 – Present · Remote

- Embed with each client from discovery to production: map the real process with owners and operators, agree on success metrics, ship the agent into their systems and iterate against real usage. Turn repeated patterns into reusable modules and a delivery playbook (AI Sprint: a working agent in three weeks).

Fullstack AI Developer · Luzid (Silicon Valley startup) Feb – Dec 2025 · Remote

- Engineered LLM features for an agentic SAP orchestration platform across three major versions: prompt versioning, regression evals, failure analysis. Started with the Brazil enterprise deployment and scaled to global rollout; the platform is used by NTT Data, delaware, Vivo and TIVIT. TypeScript, Python, AWS.

Co-founder & Community Lead · Fundación Código Abierto / GDG Barranquilla 2024 – Present · Part-time

- Built a 5,000+ developer community on Colombia's Caribbean coast: Barranqui-IA, the region's largest AI hackathon (three editions, with Google), and TechCaribe Fest (up to 2,200 attendees). Designed an AI engineering curriculum taken by 400+ students. SpaceX Ambassador for the Cursor builder community in Barranquilla since 2026.

Fullstack Technical Lead · Gorilla Logic (US clients) Oct 2020 – Sept 2023 · Remote

- Led 6-developer teams for US clients: at Flowhub, a tax-calculation engine (+60% tax accuracy); at Polco, data dashboards used by non-technical civic staff (+25% data comprehension in user testing). Code reviews, onboarding and mentoring.

Lead Developer · Urvita (Mexico, proptech) Oct 2019 – May 2024 · Remote, part-time

- Built the internal platform and mobile app that replaced spreadsheets across the operation, saving 100+ operations hours per month; introduced Scrum, doubling team throughput.

Earlier: instructor and lead instructor at Dev.F and Centraal Academy (2016 – 2019); frontend and fullstack roles at **Beek.io (Y Combinator)**, Playbusiness, and Hal.

EDUCATION & STACK

- Business Administration coursework, ITAM (Mexico City), 2013 – 2017. Software engineering program, Dev.F, 2015 – 2016.
- TypeScript, Node.js, Python, FastAPI, React, Next.js, PostgreSQL, AWS, Docker, CI/CD. LLM agents (Anthropic, OpenAI, ElevenLabs), MCP, RAG, evals, WhatsApp Cloud API. Writing: "Anatomy of a Coding Agent" (Medium).