

Sebastián García-Moreno Zinchenko

AI / Agent Engineer

Guadalajara, Mexico (GMT-6 · overlap EST & CET) · Open to Remote (US / Europe) · Open to visa sponsorship ·
segamozi@gmail.com · github.com/Sebastiangmz · LinkedIn

Summary

AI / Agent Engineer focused on shipping production-grade agentic systems. I build multi-agent pipelines, LLM orchestration layers, and RAG systems with reliability built in — defensive model routing, provider fallback chains, evaluation harnesses, and end-to-end observability. I also do harness engineering: customizing and tuning AI engineering harnesses (agent runtimes, tool layers, dev-loop scaffolding) per workflow. I design and operate the internal agent stack and showcase products of an early-stage AI studio. Open-source contributor to promptfoo (a leading LLM evaluation & red-teaming framework).

Experience

AI / Agent Engineer — Agentic Engineering Agency Feb 2026 – Present · Remote

Early-stage AI engineering studio building agentic systems for SMBs; I operate the full technical stack — research, build, deploy, observe. I lead client builds and internal tooling across multi-stage agent pipelines with defensive model routing, RAG systems, self-hosted agent frameworks, evaluation and TDD tooling, custom agent harnesses, and end-to-end observability.

AI / Agent Engineer — Voblakye May 2023 – Present · Remote

- Ongoing software & AI engineering work across internal initiatives.

AI & Automation Developer — TendencIA Dec 2024 – Mar 2026 · Remote

- Built conversational AI products: WhatsApp & Telegram bots, customer-support chatbots, and internal automations, integrating third-party LLM APIs into client-facing systems.
- **AIRE (real-estate app spun out of TendencIA)**: validated and integrated AI-driven features end-to-end, ensuring reliability of LLM-backed flows; owned most of the frontend (Flutter + Supabase).

Projects

- **PriceGenius** — AI pricing-optimization SaaS for e-commerce; multi-stage agent pipeline with defensive model routing that produces grounded recommendations.
- **Curia** — AI legal-intelligence platform: court-notification monitoring, procedural-deadline computation, and document analysis with verified citations.
- **AgentForge** — Self-hosted agent framework with OpenAI-compatible REST/SSE API, Discord / Telegram adapters, and MCP / A2A support.
- **Triage** — AI incident-triage agent: RAG over a production codebase with root-cause analysis and automated ticketing.
- **SpecSafe** — TDD framework for AI-assisted development: structured planning with spec-sliced, test-first implementation.
- **omp-pantheon** — Custom OMP-based agentic-engineering harness: specialist agents, eval-evidence gates, durable memory, trace hygiene, and bundled agentic-branch-e2e skill for branch-level E2E verification with frozen criteria and PASS/FAIL/INCONCLUSIVE verdicts. TypeScript.
- **CodeRAG** — Grounded RAG system for Q&A over code repositories, with verifiable citations; ships a dedicated MCP server (tools, resources, prompts) for agent integration. Python.

Open Source

- **promptfoo** — LLM evaluation & red-teaming framework. Merged: CLI rendering fix preventing API error logs from corrupting the progress bar (#7646); validator test-coverage suite (#7641). Additional PRs in review (Gemini retry/backoff, web UX).
- **pal-mcp-server (BeehiveInnovations)** — merged docs contribution: comprehensive WSL setup guide for Windows users (#136).

Skills

- **Languages:** TypeScript, Python, Dart, SQL.
- **AI / LLM:** Multi-agent orchestration, LLM routing & fallback chains, RAG & vector search, prompt engineering, LLM evaluation & red-teaming, MCP & A2A, voice-agent prototyping, local models (MoE, quantization), Mastra, OpenRouter. Agent harness engineering: Claude Code, ForgeCode, oh-my-pi, NousResearch Hermes, GSD (Get Shit Done), and Mastra runtimes with custom tool layers. AI observability: Langfuse, PostHog, Opik, BugSink for traces, token-cost tracking, latency monitoring, and error visibility.
- **Web / Mobile:** React, Next.js, TanStack (Router / Query / Start), Convex, Remix, Better Auth, Tailwind CSS, shadcn/ui, Flutter.
- **Data / Infra:** PostgreSQL, Supabase, Convex, LibSQL, Cloudflare (Workers / D1 / R2 / Durable Objects), Vercel, Docker / Docker Compose, Git, CI/CD.

Education

B.Eng. in Computer Systems Engineering — ITESO, Guadalajara (in progress). Spanish (native) · English (professional working proficiency — fluent).