

Carlos Galarza

Senior Full-Stack & AI Engineer | Rust, Go, TypeScript, Python

Popayan, Colombia (UTC-5) | English and Spanish | +57 322 844 5281
carloslfu@gmail.com | carlosgalarza.com | [LinkedIn](#) | [GitHub](#)

PROFILE

Software engineer shipping production applications since 2010, with four years of RAG experience and about five years leading engineering teams. Owns architecture, implementation and production operations for full-stack and AI products.

SELECTED AI ENGINEERING

Sofia, production AI assistant for email and CRM (2026): Built retrieval combining pgvector embeddings, PostgreSQL full-text search, HyDE, reciprocal-rank fusion and GPT reranking. Implemented persistent memory with linked entities, OAuth-protected MCP tools and retrieval evaluations; improved production logging reliability. TypeScript, Vercel AI SDK, OpenAI and PostgreSQL.

Stratas Assistant, Randall Reilly: Built a chat interface connecting three data products. Platform usage tripled, client adoption reached 90%, and the product contributed more than \$1 million in new ARR. Python, FastAPI, LangChain and OpenAI.

Slotstream: Built a Swift/MLX/Metal engine that streams a 125B model (about 105 GB) from SSD on a 48 GB Mac, using expert caching and custom GPU kernels. Reached #1 on Show HN; awarded Guillermo Rauch's personal \$1,000 open-source grant.

EXPERIENCE

Founder, Slotstream | Aug 2026 - Present

Released the local inference engine; building a native Mac application for private AI.

Founder, VibeCraft Inc. | Jul 2025 - May 2026

Built a platform providing dedicated computers for AI agents to use real browsers, terminals and filesystems.

Software and AI Consultant, Evermuse | May 2025 - Mar 2026

Built backend APIs, worked on model routing and an AI SDK migration, and debugged production systems using Firebase, GCP and GitHub Actions.

Independent Software Engineer and Consultant | Feb 2010 - Jul 2025

Delivered production products for US companies, including Stratas Assistant. Architecture, implementation, deployment, technical workshops and team coaching.

Founder, CodePortal | Jun 2023 - Nov 2024

Built a natural-language app builder covering the database, backend, frontend and deployment.

Tech Lead / Senior Full-Stack Engineer, Slang | Mar 2022 - May 2023

Promoted to tech lead; mentored engineers and helped resolve PostgreSQL performance bottlenecks. Ruby on Rails, React, TypeScript and Java.

Founding Software Engineer / Tech Lead, Tribe Labs | May 2020 - Feb 2022

Led four engineers. Helped migrate the live Firebase/GCP backend to Node.js and PostgreSQL on AWS. Built real-time video and mobile features with React Native and WebRTC; worked on infrastructure with Terraform.

CTO and Co-Founder, Lici/Matic | Nov 2018 - Nov 2019

Built a SaaS for finding government contracts with Go, PostgreSQL, React, Docker and GCP. Hired and led three developers.

Earlier roles: Senior Software Engineer, Globant (Feb - May 2020), Quantcast Choice; Software Architect, SchoolApply (Aug - Dec 2018), product development and XState contributions; Software Engineer, CoachCare (Feb - Aug 2018), healthcare frontend and backend systems.

SKILLS

Backend and systems: Rust ([Deno runtime example](#)), Go, Python, FastAPI, Django, Node.js, Swift, Rails, REST, GraphQL.

Frontend: TypeScript, React, Next.js, React Native, JavaScript, HTML/CSS.

AI and data: RAG, LangChain, Vercel AI SDK, MCP, Amazon Bedrock, PostgreSQL/pgvector, Redis, PyTorch.

Cloud and quality: AWS, GCP, Terraform, Docker, GitHub Actions, OpenTelemetry, Vitest, Playwright.

Education: Self-taught through software engineering and AI study and practice.