

Felipe Vargas

Backend & AI Engineer · 5 years in production

Medellín, Colombia (UTC-5, full overlap with US business hours) · Remote · Open to relocation

felipevargas.bz@gmail.com · +57 301 730 4596 ·

linkedin.com/in/felipevargas-bz · github.com/felipevargas-bz · felipevargas.tech

SUMMARY

I build and operate the systems behind AI products — agent graphs, multi-tenant serverless on AWS and GCP, and the infrastructure underneath. I also teach support sessions for 300–500 engineering students and do supercomputing research.

EXPERIENCE

Backend Lead — au2.bot

2023 – present · Remote

Conversational AI platform for paid-social campaigns. I define the architecture the team builds on.

- Built and operate the platform's multi-tenant cloud foundation on AWS — Lambda, API Gateway, DynamoDB, Step Functions, EKS — serving clients that receive 1,000+ inbound messages a day at campaign peak.
- Removed three paid SaaS dependencies from the critical path, replacing a vendor conversational platform, the messaging infrastructure and third-party integration tooling with services the company owns and operates.
- Took the AI layer through four generations — function calling, tool use, MCP, and multi-model agent graphs in LangGraph — so the product is no longer tied to one model provider.

Backend Developer — El Socio Digital / APES

2022 – 2023

- Built the warehouse system for port and bulk cargo: every load weighed at the port and again at the warehouse, shrinkage reconciled kilo by kilo against the bill of lading. The discrepancy between the two weighings is the product.
- Integrated an accounting ERP with an online storefront in both directions — a sale on either side updates the other — and integrated invoicing with the national tax authority.

Backend Developer — Viva

2021 – 2022

- Owned low fares on vivaair.com: a ± 1 -month cheapest-price calendar where every cell is a live asynchronous query against the Amadeus Navitaire API. Built on asyncio, aiohttp and FastAPI with Redis caching.

TEACHING AND RESEARCH

- UNAD — teach the official engineering support sessions: Python, FastAPI, AI and agent development, architecture and server fundamentals. Cohorts of 300–500; 380 at the most recent session. 2023 – present
- UNAD, GRUSLIN free-software research group — built the backends of SAMP, a monitoring and accreditation platform now in production at the university, and PythonLab, the code-execution environment used in the Fundamentals of Programming course.
- Supercomputing — SC25: selected for the SIGHPC/ACM Immersion Program, one of around twenty undergraduates worldwide, St. Louis. SC26: Student Volunteer, Chicago. SC24: participant.

SKILLS

Languages Python · Go · SQL · C

Cloud AWS (Lambda, API Gateway, DynamoDB, Step Functions, EKS, CloudFormation) · GCP · Terraform

AI LangGraph · LangChain · RAG · vector databases · MCP · open-source models served locally · fine-tuning

Backend FastAPI · Flask · Django · asyncio · aiohttp · Celery · WebSockets · PostgreSQL · Redis

Platform Docker · Kubernetes · CI/CD · multi-tenant architecture

EDUCATION AND HONORS

Systems Engineering — UNAD · Software Engineering Foundations — Holberton School, 2021

Devsu code-jam 2021 — Contest Finalist, ~5,000 programmers across Peru, Ecuador and Colombia