

CARLOS VEGA

Backend Engineer | LLM-Integrated Systems | Python · Go · PostgreSQL · AWS

English: Fluent (C2) | Spanish: Native | vezc.vega@gmail.com

linkedin.com/in/cvegaz | github.com/cvegaz | cvegaz.github.io

PROFESSIONAL SUMMARY

Backend engineer building production systems that put LLMs behind deterministic guardrails. Designed, built and deployed two live products — a voice-to-spreadsheet capture platform and a play-by-play sports statistics engine — on FastAPI, PostgreSQL and AWS, with over 1,000 tests that run offline, Terraform-provisioned infrastructure and credential-free continuous delivery. Also builds data infrastructure in production: Airbyte/dbt pipelines and Go REST APIs on a real estate investment platform. Fifteen years running a collegiate sports program — staff, budget and stakeholders — precede the engineering work and shape how I scope it. Looking for backend or platform roles where LLM integration is part of the product, not a demo.

TECHNICAL SKILLS

Languages: Python (FastAPI, asyncpg, pandas, openpyxl), Go (Gin, pgx), SQL, TypeScript / React (working level)

Backend & Data: PostgreSQL 16 — schema design, versioned SQL migrations with rollback, materialized views; REST API design; layered architecture (routes → services → repositories → models); dbt, Airbyte, ETL pipelines

LLM Engineering: OpenAI (gpt-4o, gpt-4o-mini, Whisper), Anthropic Claude API, MCP integration, JSON-mode structured extraction, prompt design, evaluation harnesses over golden sets

Cloud & DevOps: AWS (EC2, RDS, ECS Fargate, SSM, IAM, S3), Terraform (IaC), Docker & Compose, Caddy / Nginx, GitHub Actions — multi-arch images to GHCR, OIDC federation, SSM-based deploys

Testing & Practice: pytest, pytest-asyncio, Hypothesis, Vitest, test-first development, Architecture Decision Records, PR-driven Git flow, spec-driven development (Kiro)

Auth & Integrations: JWT / bcrypt, OAuth2 & OIDC (Keycloak, Google IdP), Stripe, Airtable

SELECTED ENGINEERING WORK

CLOKIMX | Workforce Attendance Platform (Contract, Part-time) · cloki.com.mx

2026 – Present

Backend & Full-stack Contributor — Go · PostgreSQL · React (team of three, code reviewed by the technical lead)

- Designed an immutable attendance model in which corrections stack on top of the original punch instead of overwriting it, backed by an append-only audit trail enforced through Postgres permissions rather than application convention — the log holds even when a handler is wrong.
- Derived every attendance judgment at read time: punches paired against the effective schedule (employee override, else area), shifts crossing midnight, forgotten shifts auto-closed — with no rollup tables, so a late correction resolves without a backfill.
- Closed a check-then-insert race on punch initiation with a transactional advisory lock, so a resent one-time link reuses the pending request instead of spending a second biometric verification.
- Introduced the frontend test harness the project lacked, authored half of its test files, and wrote its architecture decision log (41 entries; 27 of them decisions I made, the rest documented from existing code).

VOXA | Voice-to-spreadsheet data capture · tryvoxa.com · github.com/cvegaz/voxa

2026

Sole designer and engineer — Python/FastAPI · React · PostgreSQL · AWS

- Shipped a live system that turns spoken narration into structured spreadsheet rows: schema detected from a user-uploaded Excel template, Whisper transcription, gpt-4o-mini extraction, row appended and exported — running on AWS EC2 behind Caddy, with DNS, backups and CI/CD provisioned in Terraform.
- Contained the cost of an unauthenticated endpoint that spends real money on every call: hourly and daily per-IP limits plus a persisted budget ledger, chosen over a sign-up wall so the public demo stays frictionless.
- Delivered releases with no long-lived credentials and no inbound SSH — GitHub OIDC into AWS, multi-arch images pushed to GHCR, rollout executed through SSM Run Command.
- Kept 660 tests passing with no network and no credentials, including one that asserts the offline property itself after a leaked environment variable let the suite pass silently for weeks; recorded 19 architecture decisions with alternatives and accepted trade-offs.

Sole designer and engineer — Python/FastAPI · PostgreSQL · React · AWS

- Separated perception from decision: the model extracts only what the statistician narrated — an unmentioned field stays absent, never a courtesy zero — while a deterministic rules engine resolves down and distance, possession, turnovers, yard-line math, scoring and penalties. Narration and derived state persist in separate columns so the two never blur.
- Built an evaluation harness over a golden narration set scoring per-field accuracy and contract violations; the latest run resolved 100 of 103 expected fields across 16 plays, counting courtesy fills as a separate failure mode.
- Made post-game corrections safe: because each play's state derives from the previous one, editing a play re-runs the engine across the rest of the game and rewrites every downstream state.
- Extracted the domain-agnostic transcription and extraction pipeline into an installable package (voxa-core, 163 tests) consumed through a single prompt-builder interface, keeping product schema and prompts out of the engine.
- Shipped role-scoped JWT auth and a 15-table catalog (leagues → seasons → teams → games) on versioned SQL migrations; 412 tests run offline against the real parser and rules engine with only the model faked.

PROFESSIONAL EXPERIENCE

THEISDRIVEN / 270PARK | Real Estate Investment Platform

2026

Backend & Frontend Developer (Part-time, Remote)

- Architected and implemented an end-to-end pipeline consolidating 3 Stripe accounts and Airtable into a unified PostgreSQL warehouse via Airbyte → dbt → REST API, serving the production dashboard.
- Built production Go/Gin REST APIs for subscriptions, one-time purchases and investor–GP matching, backed by materialized views for sub-second reads from the React dashboard.
- Designed and deployed the AWS footprint with Terraform — RDS PostgreSQL, ECS Fargate for dbt runs, SSM Parameter Store — across dev and prod.
- Integrated Keycloak SSO (OAuth2 / Google IdP) with custom email-gating middleware, and authored the dbt models unifying customers, subscriptions and purchases under a 24-hour staleness contract agreed with the business.

TEC DE MONTERREY Mexico City | College & High School Activities Department

2011 – Present

Program Lead — Collegiate Football Operations & Analytics**AI-Enabled Athletic Performance Database (2025–2026)**

- Engineered an ETL pipeline turning 1,620+ athletic evaluations scattered across Excel files into a validated, normalized PostgreSQL database covering 229+ athletes.
- Integrated MCP so coaching staff query that database in natural language through Claude; currently designing a custom MCP server with curated tools, read-only roles and audit logging.

Program Leadership & Operations Automation (2011–2024)

- Ran a collegiate American football program — 16 coaching and support staff, 100 student-athletes per season, an MXN 7M (≈US\$350K) annual budget — owning logistics, payroll, scholarships and vendor contracts; as head coach through 2024, reached the national semifinal or final in five consecutive seasons.
- Cut bi-weekly nutrition consumption reporting from full workdays to 15 minutes with automated Power Query pipelines, and built Power BI dashboards over SQLite and IoT sensor feeds (WitMotion, VITRUVÉ encoders) for 400+ athletes.

PLAYPRO SYSTEMS | Video Services and Statistics

2008 – Present

Founder & Owner

- Founded and run a video and statistics service for American football, sized to a niche the market was not covering; playPro Stats is the productization of that operation's data-capture problem.

EDUCATION & CERTIFICATIONS

Master's in Software Development with AI Tools (in progress)

BIG School

Master's in Telecommunications Management

Tec de Monterrey Mexico City · Exchange: Växjö Universitet, Sweden

BSc in Electronics and Communications Engineering

Tec de Monterrey Mexico City · Exchange: Södertörns Högskola, Stockholm, Sweden

Courses: Retrieving, Processing & Visualizing Data with Python (Coursera) · Excel Skills for Data Analytics (Coursera) · Python + AI + Data Analysis · MySQL & MariaDB · Advanced Power BI · Git & GitHub · DASM · Leadership Diploma