

TIMOTHY CEVALLOS

Senior Full Stack Engineer — AI Retrieval Systems & Payments Infrastructure

+1 (239) 776-4510 • timcev@gmail.com • Remote (US / Global) • Portfolio • GitHub • LinkedIn

SUMMARY

Senior engineer with 9+ years shipping production systems end to end, including six years across two engagements with Athlean-X, a consumer fitness brand serving 100k+ daily visitors. Built and deployed the platform's Retrieval-Augmented Generation (RAG) customer support system — owning retrieval quality, hallucination control, and streaming infrastructure from architecture through production rollout — after four years maintaining the high-traffic property underneath it. Equally fluent in payments infrastructure, having shipped a multi-tenant billing engine on Stripe Connect with idempotency-protected plan changes and full audit trails.

CORE TECHNICAL SKILLS

AI / Retrieval: Retrieval-Augmented Generation (RAG) pipeline design, vector databases, Qdrant, pgvector, sentence-transformers, hybrid semantic + keyword search, RRF ranking, prompt engineering & versioning, Anthropic / OpenAI / Ollama integration, streaming inference

Backend: Python (FastAPI), NestJS, Node.js, PHP (Laravel, Zend), PostgreSQL, Redis, Prisma, WordPress VIP (enterprise), WooCommerce

Frontend: React, Next.js, TypeScript, Tailwind CSS, shadcn/ui, Framer Motion, Chart.js, Alpine.js

Payments & SaaS: Stripe Connect, subscription lifecycle, webhook handling, invoicing, payouts, multi-tenant architecture, RBAC, Clerk auth, idempotency & audit design

Infrastructure: Docker, CI/CD, AWS Lightsail, Cloudflare

Accessibility: WCAG 2.1 AA, ARIA, keyboard navigation, screen reader optimization

SELECTED WORK

AI Support Platform — Athlean-X

2025

Next.js · FastAPI · Qdrant · pgvector · Redis · PostgreSQL · Docker | delivered via Cevallos Systems, on a platform I maintained 2018–2022

- Shipped a production RAG support assistant for a fitness brand serving 100k+ daily visitors, supporting an estimated 2,500+ customer inquiries per month with a target of 65% resolution without human escalation.
- Eliminated the system's highest-severity failure mode — recommending nonexistent programs — by constraining generation to a verified catalog with category filtering and explicit fallback paths, targeting a reduction in invented-product responses from 12% to 0% across 500 evaluation cases.
- Improved retrieval relevance using hybrid dense vector search (sentence-transformers all-MiniLM-L6-v2, 384-dim) and keyword matching with Reciprocal Rank Fusion, targeting a 22% improvement against a hand-labeled query set.
- Built an ingestion pipeline designed for 250,000+ historical Zendesk tickets, incorporating spam filtering and deduplication to reduce low-signal content before embedding, with an estimated 18% filtering rate.
- Delivered token-streaming chat over Redis pub/sub, targeting 350 ms p95 time to first token, with Anthropic, OpenAI, and Ollama pluggable behind a unified model-configuration interface and lazy client initialization.
- Built an admin dashboard with prompt versioning, embeddings management, and live analytics, enabling non-engineers to tune assistant behavior without requiring a deployment.

Paganki — Multi-Tenant Invoicing & Billing SaaS

2024 – Present

Next.js · NestJS · Prisma · PostgreSQL · Stripe Connect · Clerk | independent product

- Architected the full invoice lifecycle (draft → open → paid/voided) on Stripe Connect checkout sessions, with custom payment recording and automatic status transitions.
- Designed an idempotency layer (PlatformBillingAction) that prevents duplicate charges during concurrent plan changes — the failure class most likely to produce customer-visible billing errors — with audit metadata retained for compliance and debugging.
- Built the subscription billing engine against Stripe webhooks, covering the complete lifecycle with plan-change tracking and MerchantPlanHistory audit trails.
- Implemented multi-tenant isolation with Clerk auth and RBAC, enforcing merchant-scoped data access across every API module.
- Shipped marketing and dashboard UIs in five languages (EN/ES/DE/FR/IT) on a NestJS API with typed DTOs, guards, and CI/CD.

EXPERIENCE

Cevallos Systems — Founder & Senior Full Stack Engineer

2018 – Present

Remote | Independent consultancy, run concurrently with the roles below | Clients: Athlean-X (returning), Walters Construction, Environmental Landscaping

- Delivered end-to-end systems for clients across AI, e-commerce, and enterprise — requirements, architecture, UI/UX, DevOps, testing, and production rollout, typically as the sole engineer.
- Raised client site performance by 20–40+ Lighthouse points through code audits, caching strategy, and rendering optimization.
- Removed 20+ cookie dependencies from a WooCommerce parking reservation platform by migrating to server-side WC sessions, resolving race conditions that had been producing incorrect bookings.
- Built e-commerce automation across WooCommerce and custom PHP — dynamic pricing, bundling logic, advanced checkout flows, and QR-code reservation systems.
- Shipped WCAG 2.1 AA compliant calculators and interactive components with full ARIA labeling, keyboard navigation, skip links, and screen reader support.
- Rebuilt the Walters Construction site as a high-end brand presentation, translating a luxury home builder's positioning into typography, layout, and visual storytelling.

Nielsen — Full Stack Developer

Jan 2022 – Jul 2024

Remote | Enterprise publishing platform on WordPress VIP

- Cut recurring report production from weeks to minutes by automating Google Sheets data pipelines with dynamic chart generation, replacing manual chart assembly for editorial teams.
- Partnered directly with marketing, SEO, content, and security stakeholders to scope and validate each build, translating non-technical requirements into tooling those teams adopted.
- Shipped against WordPress VIP enterprise standards — automated PHP and JavaScript coding-standard checks, mandatory code review, performance budgets, and security requirements on a high-traffic publishing environment.
- Partnered with a dedicated QA analyst through each release cycle, writing reproduction cases and running manual and cross-browser validation to keep regressions out of production.
- Built reusable, high-performance UI components in Alpine.js and Tailwind on custom build workflows, adopted across multiple editorial teams.
- Developed internal tooling for deployment, content management, and cross-team collaboration.
- Improved site performance and SEO through code audits, accessibility fixes, and script restructuring.

RCH Tech Solutions — Full Stack Developer

2018 – 2022

Remote | Primary developer on Athlean-X, the agency's largest account — 100k+ daily visitors

- Owned day-to-day development on Athlean-X across four years — WordPress themes, plugins, and PHP APIs serving 100k+ daily visitors against a large program and supplement catalog.
- Led a full PWA redesign of the property that improved mobile performance, accessibility, and user retention.
- Built a Laravel-based in-browser video editor that automated rendering and cut production turnaround time for a content operation built on video.
- Implemented custom dashboards, analytics modules, and workflow automation supporting the editorial and marketing teams.
- Developed features end to end for a custom user portal on Zend Framework MVC — data layer through front end — serving thousands of concurrent users across international markets.

EDUCATION

Flatiron School — Full Stack Web Development Certificate