

Azahid García

Applied AI Engineer | LLM Evaluation · Agentic Systems · MCP | Python · TypeScript

Mexico City, MX (open to relocation · US ET overlap) · azahid.garcia@gmail.com · +52 5567823655

linkedin.com/in/azahidgarcia · github.com/AzahidGarcia · strivark.com · portfolio-azahid-garcia.vercel.app

SUMMARY

Applied AI Engineer specializing in LLM evaluation, agentic workflows, and multi-agent orchestration in Python and TypeScript/Next.js. I design eval loops, adversarial benchmark tasks, and structured output contracts that make model quality measurable — and quality degradation visible before it reaches users. At Outlier I design adversarial evaluation tasks that stress-test frontier AI models (Terminal Bench 3.0); at Strivark, the B2B AI platform I founded, I ship LLM products to paying enterprise clients. Four years of enterprise consulting (WMS/ERP across LATAM and EMEA) means I sit comfortably at the boundary between an ambiguous business problem and the engineering decision that resolves it.

TECHNICAL SKILLS

LLM Evaluation & Agents: Adversarial benchmark design, rubric and verifier design, deduction-based scoring, eval loop instrumentation, prompt architecture (system prompts, skills, sub-agents), Claude API, OpenAI SDK, Ollama, MCP (Model Context Protocol), Pydantic AI, FallbackModel orchestration, tool schema design, structured output contracts

Backend & Data: Python (primary), PostgreSQL (RLS, partitioning, schema design), Supabase (Edge Functions, Auth, Realtime), pgvector for RAG, n8n automation pipelines, REST API design, ETL/ELT

Frontend & Full-Stack: TypeScript/JavaScript (ES modules), Next.js 14, React, Tailwind CSS, Radix UI, GSAP/Framer Motion — production UIs at scale

Agentic Workflows: Multi-agent orchestration, tool registration and discovery, guardrails, output quality gates, fallback chain design, per-channel success/failure instrumentation

Infrastructure: Docker, Vercel (production), GitHub Actions (CI/CD), Azure (Fundamentals certified), Git

Enterprise Context: Blue Yonder WMS/ERP (dual Technical + Functional certification), cross-regional LATAM/EMEA async delivery, technical communication with non-technical stakeholders

KEY PROJECTS

Busqueda_Empleo — LLM-Scored Job Pipeline · [Eval Calibration Case Study](#) *Apr 2026 – Present*

- **Eval integrity recovery:** diagnosed an LLM scorer producing 93% fit confidence on a role whose honest fit was ~55%; root cause was an additive rubric structurally biased toward optimism. Redesigned to deduction-based scoring (explicit penalties per missing hard requirement, seniority-gap formula, pre-LLM title filters), eliminating all confirmed false positives in the following validation run
- **End-to-end pipeline:** 9 job boards scraped → deduplication → heuristic pre-filter → LLM deduction scoring → storage → automated delivery; 300+ roles scored across two tracked channels, thresholds calibrated against a hand-labeled ground-truth set

OpenClaw — MCP Agent-to-Tool Bridge · [Personal Project](#)

Jan 2026 – Present

- **Schema-driven tool infrastructure:** socket-based MCP bridge exposing 15 tool categories to AI agents, with tool registration, discovery protocol, and schema validation
- **Eval-first instrumentation:** tracked per-tool success/failure and response time across 100+ agent interactions; every systematic tool misuse mapped to ambiguous schema descriptions, not broken code — an insight that reshaped how I write tool contracts, system prompts, and structured output specs

Revenue Engine — Multi-Agent Orchestration · [Personal Project](#)

Mar 2026 – Present

- **Resilient LLM infrastructure:** Pydantic AI FallbackModel chain (Ollama → Groq → Gemini) across 6 parallel agent channels; per-model latency and fallback frequency instrumented so degradation is visible before it affects output
- **Structured output as quality gate:** Pydantic schemas fail loudly on malformed model output; failed outputs logged with triggering input for prompt iteration — never silently discarded

Strivark — B2B AI Consulting Platform · [Founder](#) · [Lead Engineer](#)

Jan 2026 – Present

- **Production architecture:** Next.js 14 + Supabase (PostgreSQL with RLS multi-tenant isolation, Auth, Edge Functions, Realtime) + Stripe + SPEI + Mifiel + n8n — from zero to paying enterprise clients

- **AI observability in production:** Claude API workflows with structured failure alerts and output logging, surfacing degraded agent behavior before clients notice

img2blend — Python CLI Developer Tool · [Personal Project](#)

Nov 2025 – Present

- **Pipeline reliability:** image → Blender conversion via OpenCV/PIL → potrace/vtracer → SVG → Blender subprocess; each stage fails explicitly — no silent corruption, no partial renders downstream

WORK EXPERIENCE

Founder & AI Engineering Lead · [Strivark](#)

Jan 2026 – Present

- **Shipped production AI features:** designed and deployed LLM-powered agent workflows for enterprise clients using Claude API; architected system prompts and multi-turn orchestration for complete client-facing product suites
- **Prompt iteration pipeline:** real-user evaluation and output scoring across active client engagements — quality measured against concrete criteria, not synthetic benchmarks

AI Training Specialist (Freelance) · [Outlier](#)

Sep 2025 – Present

- **Adversarial evaluation:** design PhD-level verifiable tasks that stress-test reasoning limits in frontier AI models (Terminal Bench 3.0), spanning data engineering, Python, quantitative, and mathematical domains
- **Benchmark task architecture:** built a reusable task-creation framework (spec parsing → verifiable environments → final verifiers) improving consistency and coverage across technical evaluation domains

Senior Software Consultant · [Netlogistik by Argano — LATAM + EMEA](#)

Jan 2025 – Sep 2025

- **Embedded cross-regional:** worked with client engineering teams across LATAM and EMEA on enterprise system implementation and automation; collaborated on ML projects for warehouse optimization impacting multi-regional teams
- **Async by default:** operated across UTC-6/EMEA continuously — consistent overlap with US and European business hours

Software Consultant · [Netlogistik by Argano — LATAM + EMEA](#)

Jun 2022 – Dec 2024

- **Enterprise automation:** automated production WMS/ERP environments with Python and SQL; implemented Blue Yonder WMS/ERP systems, earning dual Technical and Functional Consultant certifications
- **Cross-functional delivery:** managed multiple concurrent client projects across LATAM and EMEA, translating complex technical behaviors into actionable specifications for non-technical stakeholders

Junior Developer · [DocSolutions](#)

Jan 2022 – Jun 2022

- **Backend and ML foundations:** backend scripting, SQL optimization, enterprise system troubleshooting, and machine learning training pipelines during cloud platform onboarding

EDUCATION

B.Sc. Systems Engineering (in progress): TecNM — Tecnológico Nacional de México · 2021 – Present

Partial B.Sc. Biological Chemistry: UNAM · 2019 – 2021

Technical Degree in Programming: CECyTEM Tultitlán · 2015 – 2018 — highest academic achievement of the generation, completed alongside upper secondary education

CERTIFICATIONS & CONTINUOUS LEARNING

- C1 Cambridge English Advanced (2023) · Blue Yonder WMS Technical + Functional Consultant, dual certification (2025)
- Azure Fundamentals — Microsoft LaunchX (2021) · AI Fundamentals — Microsoft Innovation Mexico (2021) · Java Full Stack — Generation México (2021)

LANGUAGES

Spanish: Native (Mexican Spanish, CDMX)

English: C1 Advanced (Cambridge certified) — presenting technical work to engineering and non-technical audiences across LATAM and EMEA