

Gustavo Schneider

AI Infrastructure Engineer · Rust · Distributed Systems

São Paulo, Brazil · gmhschn@gmail.com · github.com/gmhelmold · linkedin.com/in/gustavo-schneider

SUMMARY

Software engineer focused on AI infrastructure, distributed systems, and Rust — runtimes, execution systems, content-addressed storage, and tooling for AI agents. I build systems that are structurally incapable of flattering themselves: instrumentation that publishes what it failed to explain, documentation that fails CI when it drifts from the code it cites.

EXPERIENCE

Independent Engineer — HuGR (personal infrastructure project)

Solo · remote · March 2026 – Present

A personal project, not a company: no external customers or third-party traffic. Built solo by orchestrating fleets of LLM coding agents through a harness I wrote, gated on checks that verify their work; the architecture, contracts, and invariants are mine.

CoreLink — multi-tenant content-addressable cache and CI acceleration fabric · closed source; walkthrough on request

Cloudflare Workers, Durable Objects, Firecracker, R2, D1 · ~73 Rust crates + TypeScript edge · deployed across 5 Cloudflare regions

- Cut warm cache-hit latency from **~750 ms to ~44 ms** by moving the read path into the edge runtime at the caller's colo; validated the Rust→TypeScript port byte-for-byte with cross-language test vectors against live production objects.
- Designed multi-tenant storage isolation as a cryptographic property of the object address, using per-tenant HMAC-SHA256-derived key prefixes.
- Built cross-tenant deduplication of public artifacts, with revocation enforced atomically in the lookup statement and BLAKE3 re-verification on every read.
- Implemented exact usage metering on Durable Objects with region-sharded counters and 1,000-request lease blocks, cutting strongly-consistent coordinator round trips to ~0.1% of requests at zero over-serve; released shadow-first.
- Built an ephemeral Firecracker CI runner fabric co-located with the content store: per-job spawn, zero idle cost, credentials resolved by introspecting the presented token.
- Shipped native protocol surfaces — Bazel REAPI v2, Turborepo, sccache/WebDAV, OCI, Homebrew, PyPI — so a client integrates by pointing existing tooling at an endpoint, with no SDK.
- Delivered EU data residency in jurisdiction-pinned buckets, cross-region GDPR erasure with per-region attestation, and TLA+ models of the protocol invariants (44 of 47 specifications model-checked in CI).

Freelance Developer

Independent · 2025

Moved into engineering full-time in 2025. Client work in general software development, and LLM projects alongside it: agent orchestration, knowledge graphs, guardrails, and chatbots wired into systems already running — FastAPI services in Python, TypeScript on the front end.

Product Owner → Project Manager — OpenSolo Soluções

Remote · 2022 – 2024

- Founding Product Owner from day zero: led a B2B commodities marketplace for Latin America (catalog, orders, logistics, supplier onboarding) from the first stakeholder conversations through specification, development, and delivery.
- Designed and delivered B2B foreign-exchange modules integrated with Banco Bexs, acting as the primary liaison between the bank and OpenSolo across the entire integration lifecycle.
- Promoted to Project Manager: roadmap, delivery, and risk management across concurrent workstreams, in a trilingual environment (Portuguese, English, Spanish).

Project Manager — Temple Roofing

Melbourne, Australia · 2019 – 2022

Ran operations for a construction company: payments, purchasing, inventory, and team scheduling.

OPEN SOURCE CONTRIBUTIONS

Merged

- **Apache DataFusion** — github.com/apache/datafusion/pull/23484: support co-partitioned range right-side equi hash joins.
 - **Apache Iceberg (iceberg-rust)** — github.com/apache/iceberg-rust/pull/2825: add the `history` metadata table.
 - **ComfyUI** — github.com/Comfy-Org/ComfyUI/pull/14885: fix the `SaveVideo` node description, which documented images while saving video.
-

WORK IN PROGRESS

Relay — Python · POSIX shell · experimental · private; walkthrough on request. A verified-execution state machine for LLM agents: nested macro-states and sub-states, forward-only, each with its own prompt, tools, skills, and gate, declared in a profile file rather than in code. A state closes only when its gate passes — named controls that are ordinary shell commands, run outside the agent's process. A ledger and notary capture each state separately: verdicts, transcript and tool calls, and what the state cost in tokens, turns, seconds, and retries, all inside the hashed body, so a spend figure is as tamper-evident as a verdict. `relay verify` re-walks the chain as a separate program, and an LLM grader never closes a gate on its own. Its central thesis shipped with a pre-registered kill condition and went default-off when the measurement refuted it.

Lightr — Rust · open source · experimental · github.com/gmhelmold/hu-gr-lightr. Daemonless, imageless container runtime (~90k LOC): rootless cold start at **30.8 ms** on public CI, memoized rebuilds, a Kubernetes CRI backend passing **34/34 critest v1.33**, and a hand-written seccomp→cBPF compiler with a simulator test-bench.

Datarail — Rust · open source · experimental · github.com/gmhelmold/HuGR_Datarail. Kafka wire-protocol-compatible broker where storage never sees plaintext (`#![forbid(unsafe_code)]`): AEAD with Ed25519 and X25519, offline-verifiable delivery receipts, exactly-once semantics tested against a real PostgreSQL 16, plus an independent adversarial audit.

HuGR Tools — Cloudflare Workers · open source · experimental · github.com/gmhelmold/HuGR_Tools. An experiment in generating tool servers instead of writing them: a pipeline ingests a provider's API documentation (OpenAPI specs up to 8.4 MB) and emits a working MCP server — it produced **630 tools across 37 providers** (AWS, Cloudflare, Stripe, Datadog, and more). The generator covers the mechanical surface; multi-step orchestration is hand-written on top. Each provider is an independent Worker on a shared kernel supplying retry, circuit breaking, token-introspection auth, per-request BYOK credentials, and rate limiting. **8,538 tests**, canary deploys with rollback.

SKILLS & EDUCATION

- **Languages:** Rust, Python, TypeScript, Go, SQL, Bash
- **AI & LLMs:** agent orchestration (LangGraph, DSPy, Pydantic AI, Temporal), knowledge graphs (Graphiti, Neo4j), guardrails (NeMo Guardrails, LLM Guard), evaluation (RAGAS, DeepEval, LLM-as-judge), vector search (Qdrant), vLLM, Ollama
- **Agent infrastructure:** sandboxing (seccomp/cBPF, Firecracker microVMs), secrets brokering, ephemeral compute, memoized execution, MCP
- **Systems & storage:** Cloudflare (Workers, Durable Objects, R2, D1), PostgreSQL/Neon, replication and failover; content-addressed storage, BLAKE3, FastCDC, OCI, Kubernetes CRI, Bazel REAPI v2, sccache
- **Security & correctness:** Ed25519, Argon2id, AES-GCM, envelope encryption, BYOK, SBOM, code signing, SLSA; TLA+, property-based testing, fuzzing, mutation testing
- **Spoken languages:** Portuguese (native) · English (fluent) · Spanish (advanced)
- **Education:** self-taught engineer. Business Diploma (II), Greystone College, Melbourne 2019–2020; Advertising (Mackenzie) and Economics (FAAP), both incomplete