

Gabriel Fialho

Software Engineer · AI Systems, Developer Tooling, TypeScript & Python

Belo Horizonte, Brazil (UTC-3) · Remote · gabriellmmfialho@hotmail.com · [LinkedIn](#) · [GitHub](#) · gabriel.fialho.com

SUMMARY

Software engineer at SYDLE since 2023, on the AI team since March 2026, building the retrieval and provider layer behind an LLM product: hybrid BM25 and vector search fused by a Reciprocal Rank Fusion layer I wrote in JavaScript, and one facade that put eleven providers behind a single interface. Before that, nearly three years on the company's education platforms, where custom static analysis I wrote replaced hand review in a large legacy migration. Outside work I ship developer tooling: a CLI published on crates.io and npm, a published Claude Code plugin, and a content pipeline that runs a live site for \$0.06 an article.

SKILLS

Languages	TypeScript, JavaScript, Python, C++, SQL
Frameworks & Runtime	Node.js, Next.js, React, Express, Stencil.js, Web Components, Zod
Data & Storage	PostgreSQL, MongoDB, Elasticsearch, Prisma, REST API design
AI Systems	LLM integration in production, RAG and hybrid retrieval (BM25, kNN, RRF), Embeddings and vector search, Tool calling, MCP (Model Context Protocol), Provider-agnostic abstraction, Prompt engineering, Deterministic quality gates, Cost control
AI providers worked with	OpenAI, Anthropic, Amazon Bedrock, Google AI, Gemini Enterprise, xAI, DeepSeek, Perplexity, Voyage AI, AssemblyAI, LiteLLM
Architecture	Modular architecture, Hexagonal architecture, Facade pattern, Context-based front-end design
Practices	CI/CD (GitHub Actions, GitLab), Testing (pytest, Jest), Integration testing with recorded requests, Static analysis & custom lint rules, Code review, Agile/Scrum, Technical documentation (DIVIO)

EXPERIENCE

Software Engineer, AI team, SYDLE Belo Horizonte, Brazil	Mar 2026 – present
- Rebuilt the RAG retrieval layer as a hybrid of Elasticsearch BM25 and kNN vector search, fused by a Reciprocal Rank Fusion implementation written in JavaScript, fixing the exact-match queries customers reported as failing and delivering a ranking capability that was otherwise gated behind a paid Elasticsearch tier. - Unified eleven AI provider gateways (OpenAI, Anthropic, Amazon Bedrock, Google AI, Gemini Enterprise, xAI, DeepSeek, Perplexity, Voyage AI, AssemblyAI and LiteLLM) behind a single AI Connector facade, removing duplicated request logic from around 20 agent configuration classes and consolidating tool calling across 4 agent types into one implementation. - Restructured the AI product documentation into the four DIVIO quadrants and made it a required part of every merge request, so docs stay current with the code instead of decaying between releases.	

Belo Horizonte, Brazil

- Cut an estimated 150+ developer hours from a large-scale legacy migration by writing custom ESLint rules and polyfills that statically analysed 140+ classes and 300+ methods, recreating the framework's runtime context so linting worked inside a non-standard execution environment; other teams adopted the rules and they stayed on after the migration, moving a whole class of defect from runtime to edit time.
- Integrated an LLM into the discussion forum to suggest answers to support questions automatically; built a daily job that collects development-environment error logs, opens a tracking object for each and rotates ownership across the team on a round-robin roster, so every error has a named owner that day.
- Proposed and implemented a front-end architecture centralising context at the test level instead of cascading it through the component tree, making components modularly extensible; decoupled the assessment and learning systems behind a Facade.

PROJECTS

tecnoversus.com, Sole engineer

2026 – present

- Built a five-stage, git-native content pipeline that researches, writes, gates and publishes to a live WordPress site for ~\$0.06 per article, running on 8 CI/CD workflows with one human review step and no publishing credentials held by the generator.
- Designed a deterministic, LLM-free quality gate pinning 14 production invariants with tests, eliminating eight classes of content defect at zero marginal cost per run.
- Diagnosed an intermittent CI publishing failure as IPv6 resolution on IPv4-only runners and fixed it at the DNS layer, after a socket-level fix proved insufficient in production.

agentlink, Author, open source

2026

- Published an open-source CLI on crates.io and npm that lets six AI coding agents share one set of rules and skills through inode-level linking instead of generated copies, removing the drift, stale-file and orphan problems that copy-based tools create.

architect, Author, open source

2026

- Built and published an open-source Claude Code plugin, installable from a marketplace, that audits any repository and installs the engineering guardrails AI-assisted work needs: branching, database isolation, CI/CD, versioning, testing and enforcement.

Lavanderia Studio 733, Sole engineer

2026

- Shipped a booking system for a residential building's shared laundry, around 20 residents and 3 machines, on Next.js and Prisma; dropped authentication entirely in favour of apartment-number identity, removing signup friction for a non-technical audience where the cost of impersonation is effectively zero, and added web push before the start and end of each slot. Private deployment for one building, so there is no public link.

Brazilian postgraduate data mining, Two-person university project

UFMG, 2025

- Classified and clustered 1.24M Brazilian postgraduate theses across 175 knowledge areas: built the keyword cleaning pipeline (entropy and conditional-probability cuts), word2vec representations and an MLP, then K-means over the learned vectors, which recovered 11 semantically coherent groups. Report published.

EDUCATION

B.Sc. in Computer Science, Federal University of Minas Gerais (UFMG)

Mar 2020 – Jun 2025

LANGUAGES

Portuguese: Native · English: Professional working proficiency