

Thiago Polati

Senior AI Engineer · LLM Apps · Agents · RAG · Backend (TypeScript / Node / Python / Postgres)

Location: Madrid, Spain

Email: thiago@tp-coder.dev

LinkedIn: linkedin.com/in/tpolatti

GitHub: github.com/tp-coder

Website: tp-coder.dev

Professional Summary

Senior AI engineer focused on LLM applications — agents, RAG, tool-use — backed by 20+ years of engineering experience across backend systems, architecture, and delivery in business-critical environments.

I operate sophisticated LLM-application infrastructure day-to-day: a governed multi-agent collaboration system built on Claude Code, with specialised subagents, persistent file-based memory, gated batch protocols, and structured human approval points. I treat AI work as a coordination problem, not an autonomy problem. The reliability comes from refusals, evidence grounding, and review at meaningful seams — not from giving the model more rope.

My backend depth (TypeScript / Node / Python / PostgreSQL, multi-tenant systems, modernisation in messy codebases) is supporting evidence for the AI work, not a separate identity. I am strongest in applied LLM systems where governance, structured outputs, and evidence-grounded behaviour matter more than autonomy hype — with RAG and evaluation tooling already shipped and operating (a labelled eval suite, retrieval metrics, a facts-and-gates classifier), on top of shipped agent and multi-agent work.

Core Strengths

- **LLM applications engineering** — agents, RAG, tool-use, structured outputs, evaluation discipline
- **Multi-agent orchestration and governed AI collaboration** — coordinator + specialist pattern, gated batch protocols, file-based memory, human approval at meaningful points
- **Multi-tenant backend architecture** — TypeScript / Node / Python / PostgreSQL, Drizzle ORM, schema-migration discipline
- **Modernisation in messy environments** — pivots, legacy stabilisation, system-boundary repair without rewrite theatre

- **Operational discipline in business-critical systems** — verification over autonomy, refusals as design choices, audit trails over vibes
-

Selected Writing

Technical writing on LLM reliability, evaluation, and governed agent design — the thinking behind the systems above. Full set at tp-coder.dev/writing.

- **temperature=0 Is Not a Promise** — why LLM determinism is a myth, and how self-consistency voting recovers reliability. tp-coder.dev/writing/temperature-zero-is-not-deterministic
 - **Not All Nondeterminism Is Temperature** — separating prompt contradiction from sampling noise, and why facts belong in the model while decisions belong in deterministic code. tp-coder.dev/writing/not-all-nondeterminism-is-temperature
 - **Build the Eval Before the Feature** — treating the eval as a decision instrument, not an afterthought. tp-coder.dev/writing/build-the-eval-before-the-feature
 - **When Not to Reach for LangGraph** — judgment on agent frameworks vs deterministic workflows. tp-coder.dev/writing/when-not-to-reach-for-langgraph
 - **From One Generalist Agent to a Governed AI Collaboration Model** — the coordinator-plus-specialist multi-agent pattern in practice. tp-coder.dev/writing/from-one-generalist-agent-to-a-governed-ai-collaboration-model
-

Professional Experience

Technical PM / AI-Readiness Consultant

Radisson Hotel Group — via 2P Connect | March 2026 - Present

Consulting engagement on Radisson's GEO (Generative Engine Optimization) and AI-readiness roadmap for an enterprise hospitality content platform on the Tridion / RWS stack. Technical coordination and architecture across distributed teams (Spain, Argentina) in a layered enterprise organisation.

- Drove phase separation between immediate enablement (PoC / pilot fixes), foundational web work (SSR, semantic HTML, structured data, schema completeness), and advanced AI-readiness (MCP / WebMCP, machine-readable interfaces) — protecting estimation baselines from scope contamination
- Established the technical baseline for AI-consumable content: `llms.txt`, `AI-crawler robots.txt` review, Hotel / LodgingBusiness schema, FAQPage schema on key pages, Organization + sameAs, IndexNow, Wikidata entity authority
- Reviewed and pressure-tested vendor proposals for AI content workflow productionisation (Fonto + Tridion-based governed content pipeline), including cost / scope / operational ownership framing for executive review

- Translating roadmap items into honest engineering scope — naming where “AI work” is really foundational web platform work, and where it is genuinely new architecture
-

Founder · Senior Engineer / Tech Lead

2P Connect | May 2017 - Present

My own consultancy. Backend and AI-engineering contracts, plus my own products. Recent work has shifted decisively toward LLM applications, multi-agent orchestration, and multi-tenant content systems.

- **Building Draft Foundry** — a multi-tenant content platform for B2B real estate and hospitality, in active development. Designed and shipped the object-model law, the verticalization seam, hard brand/property isolation, image and publication boundaries, and the governed multi-agent operating system used to ship it. Stack: PostgreSQL 17, Drizzle ORM, Next.js 16, React 19, TypeScript, S3-compatible storage (DigitalOcean Spaces), Caddy, Docker. Live development at draftfoundry.ai · case study at tp-coder.dev/projects/draft-foundry
- **Operating a multi-agent Claude orchestration system** day-to-day — specialised subagents (engineering, product/UX, editorial, coordination), persistent file-based memory, auto-loaded rules, gated batch protocol with explicit human approval at every batch boundary. This is real LLM-applications operating experience, not a prompt collection. Companion article: tp-coder.dev/writing/from-one-generalist-agent-to-a-governed-ai-collaboration-model
- **JD Analyzer — shipped production LLM app** (tp-coder.dev/jd-analyzer) — a live TypeScript / Next.js pipeline that turns a job description into a grounded fit verdict. The architecture keeps the model on a short leash: the LLM only *extracts facts* (grounded yes/no answers), and deterministic code applies the decision logic — facts in the model, gates in the code. Reliability comes from three mechanisms, not from trusting a single call: self-consistency voting at the fact level (5 runs), a contiguous-quote grounding guard that only counts a fact if its evidence is a real contiguous span from the source — so a hallucinated fact can’t flip a decision — and a labelled eval suite (7 cases) asserting set-equality with no false positives, validated stable across repeated runs. Production-hardened (per-IP + global rate limiting, model-spend caps, graceful degradation) and built deliberately to *refuse* — no auto-apply, no automated outreach. A self-hostable open-source core (n8n, MIT) exists as a secondary deployment path. Companion article on why the refusals are the design: tp-coder.dev/writing/building-ai-tools-that-refuse · github.com/tp-coder/n8n-jd-analyzer
- **ai-eng-prep — public RAG / retrieval-evaluation repo** (github.com/tp-coder/ai-eng-prep) — a focused, public practice and portfolio repo: a Python RAG pipeline over a Qdrant vector store (chunking, embeddings, cosine retrieval with grounded refusal) plus a retrieval-evaluation harness computing MRR / Hit@1 / Hit@k.

Direct, inspectable proof for RAG, vector databases, and retrieval evaluation — small and deliberate, not a production system

- **Backend consulting and contract engagements** — multi-tenant systems, modernisation of legacy platforms, CI/CD repair, distributed-team delivery across Europe, the US, Brazil, and India. Teams from 5 to 14 engineers and QA. Regulated and business-critical environments where reliability and operational discipline mattered more than novelty

Head of Software Engineering

Edusogno / EasyPeasy Fluent | Feb 2025 - Mar 2026

Led engineering for a distributed team of 10 across student, tutor, admin, and mobile platforms in a live EdTech environment.

- **Introduced AI-assisted development practices as an engineering productivity layer** — structured use of LLM tools across spec clarification, implementation, review, and documentation. This was the early operational ground for what later became the multi-agent system at 2P Connect
- Designed a backend modernisation strategy starting from a microservices direction, then **deliberately reversed course back to monolith** when the complexity cost did not match the team's actual constraints. The credible move was to delete the dead direction, not ship it
- Redesigned and shipped a **WhatsApp messaging orchestration system** for lead and lifecycle automation — case study at tp-coder.dev/projects/messaging-orchestration-system, with companion article *The Boundary Was the Bug*
- Designed and implemented an **outgoing payments system** integrating Revolut-synced expenses with internally managed manual payments — case study at tp-coder.dev/projects/outgoing-payments-system
- Improved platform robustness, deployment processes, and developer workflows

Stack: PHP, Laravel, React, Python, AWS Lambda, AWS Step Functions, MariaDB, Docker

Earlier Career Highlights

BRQ USA · CINQ Technologies / CI&T · Mirum Agency · ExxonMobil · Wipro · Telefônica Brasil | 2008 - 2023

Progressively responsible roles across engineering delivery, technical project management, and enterprise support — the foundation that the AI-engineering and platform work sits on.

- **Senior Sales Engineering Manager, BRQ USA (2022–2023)** — led a client-facing engineering team of 5, helped secure new revenue, supported strategic US-market partnership development. Stack: Java, Angular, React, Python, Node.js

- **Global Engagements Manager, CI&T (2019–2022)** — managed a \$3.2M/year global account and coordinated delivery across teams of up to 67 people across multiple countries. Operational discipline at scale: clearer collaboration structures, Jira / Confluence rhythms, cross-team coordination
- **Software Development and IT Project Manager, CINQ Technologies (2018–2019)** — allocated at Rumo Logística, leading distributed development teams across multiple cities and partner organisations on rail-safety and industrial software in a regulated environment
- **Software Development and IT Project Manager, Mirum Agency Brasil (2017)** — managed development and distributed QA across Brazil and Canada for a UNICEF web platform program. Contributed to the launch of 190+ UNICEF country, regional, and global sites
- **Earlier roles (2008–2017)** — application support and performance work for pricing and reporting systems at ExxonMobil; Microsoft / Active Directory infrastructure support at Wipro (subject-matter expert for ADDS); project leadership at Telefônica Brasil across commercial, finance, and city-expansion initiatives

These roles built the foundation for systems thinking, operational discipline, distributed-team coordination, and delivery under real business constraints — the substrate that the LLM-applications work is built on.

Education

MBA, Project Management — ESIC Business & Marketing School (2014–2016) **BSc, Computer Engineering** — Universidade Positivo (2004–2008)

Languages

Portuguese (native) · English (fluent) · Italian (fluent) · Spanish (advanced) · French (intermediate)