

WAGNER SOARES ANDRADE DE LIMA

Generative AI Solutions Architect · LLMOps · Founder, LOGOS

São Paulo, SP, Brazil · +55 83 98831-0304 · hello@wagnerlima.cc
linkedin.com/in/wagner-lima-ai · wagnerlima.cc · thelogosai.com.br

PROFESSIONAL SUMMARY

Generative AI Solutions Architect with 16 years combining strategic communication and technical execution. I build LLM systems and business systems end to end, from concept to operation in production, with a focus on governance, observability, cost, and security. I deliver with handover: the client operates autonomously and the consultancy maintains infrastructure, security, and evolution. Cross-functional work across architecture, product, and business, translating complex demands into scalable solutions. Technical decisions based on evidence, not hype. Solid experience in advanced prompt engineering, integration with LLM APIs, MCP (Model Context Protocol) infrastructure, and driving coding agents through milestone-based builds with technical and security review.

PROFESSIONAL EXPERIENCE

LOGOS · Solutions Architect & Founder

Jan 2026 – Present · São Paulo, SP, Brazil

Digital solutions consultancy with generative AI in production. End-to-end work: discovery, architecture, implementation, deployment, observability, and handover so the client's team can operate autonomously.

- B2B management system for a bakery in real operation: NestJS + Prisma + PostgreSQL + Next.js on Render. Orders with an auditable status model, client portal, billing reports (summary and detailed, with per-product consolidation for invoicing (nota fiscal)), thermal printing, hierarchical RBAC, CI, and dependency security scanning. Replaced spreadsheets and paper notepads in the operation. Handover through hands-on training of 6 team members; the system has been running for over two months without intervention, with a declared autonomy boundary (client operates; LOGOS maintains infrastructure, security, and evolution). Living architecture documentation generated from the code, with per-session handoffs. Product decision: iFood-style status flow pulled back behind a feature flag when the operation did not sustain it, without discarding code. Project started in Feb 2025 as independent work and incorporated into LOGOS in Jan 2026.
- Multi-turn conversational intake system in production: Cloudflare Workers + TypeScript + D1 + Anthropic API + Meta WhatsApp Business Cloud API. FSM pipeline with auditable states, technical brief generation via LLM, Chatwoot integration as a read-only history panel, routing via Telegram. Operational at +55 11 5286-8025.
- LOGOS client portal and CRM, live since Aug 2026: Next.js on Cloudflare Workers (OpenNext) with D1, R2, and KV; Auth.js with client and prospect activation, rate limiting, and TOTP on the admin; HMAC webhook integrated with the intake bot; transactional emails (Resend) and retention cron; AI substrate with per-project context. Tests with Cloudflare's Vitest plugin and visual auditing with Playwright as a permanent deployment gate. Build driven by a coding agent across milestones M0–M10 with a review checkpoint between each, including a security fix (secret read at runtime, never embedded in the bundle) and post-deployment incident resolution (PBKDF2 iteration limit on Workers). Production runbook, operations manual, and handoff documented.
- TA Analyst, crypto technical analysis web app in production: Next.js on Vercel, PostgreSQL (Neon), Anthropic API with prompt caching, public Binance data. Deterministic multi-timeframe calculation without LLM; the AI interprets what the code calculated and never generates numbers. Explicit product positioning: structured analyst, not a signal generator, with invalidation before zones and an auditable history. Born from a real intake from the LOGOS bot, closing the funnel end to end.
- LLM cost and quality optimization: prompt caching on an invariant block (Anthropic), ~45% compression of the system prompt (~26K → ~14K tokens) with no loss of quality, turn cap and intake depth calibration based on iteration with production.

- Root cause analysis and structural fix in a production incident: timeout in brief generation resolved via Service Binding self-fetch between Workers, isolating intensive processing from the main worker.
- Proprietary Discovery Framework (v1.3): technical-strategic client diagnosis with a Node.js + docx + LibreOffice pipeline for PDF report generation. Delivered cases: Felipe Asafe/Ademicon, Arena Fort3 Beach Sports.
- MCP (Model Context Protocol) infrastructure: building and maintaining MCP servers for orchestrating flows with Anthropic models, including Memory Cloud (Go + SQLite) for distributed multi-project knowledge graphs.
- Pimp My Prompt: public tool (thelogosai.com.br/pimpmyprompt) as a showcase of automation and prompt engineering capabilities.

Stack: TypeScript, Node.js, NestJS, Next.js, Prisma, Drizzle, Auth.js, PostgreSQL, Cloudflare Workers/D1/R2/KV, Vercel, Render, Neon, Anthropic API, Meta WABA, MCP, Go, SQLite, Astro, Tailwind, Vitest, Playwright, Resend.

SMN Tecnologia da Informação · **AI Solutions Architect**

Jan 2024 – Feb 2026 · João Pessoa, PB, Brazil

Architecture and development of generative AI solutions for clients across multiple industries, from concept to delivery in production, at the intersection of business and technology.

- Autovoxx (real-time sales copilot with behavioral AI): evolved from a technical pitch to a partnership with investors. Real-time inference architecture during sales calls.
- Intelligent processing pipeline for tax documents: 2,300+ documents/month in multiple formats, with structured extraction via LLM and automated validation.
- Sales lead automation: response time reduced from 2–4 hours to under 5 minutes.
- Regulatory chatbot: navigation across a base of 68 regulations (~770 pages) with a 99% reduction in lookup time. Contextual retrieval architecture applied to unstructured data.
- .NET Track: 22-week technical training program (27% faster than the market standard) with a 5-stage assessment system, individual tests customized based on identified gaps, and automated grading via Microsoft Forms. Standardization of best practices and code review among interns.
- Corporate website with enterprise architecture: Lighthouse 100/100/100/100, LCP <200ms, zero regressions in production.
- Exclusive talk on strategic AI for LIDE Paraíba: 12 business leaders, technical → executive translation. Generated a contracted consulting project.

Stack: .NET/C#, Node.js, TypeScript, Azure, LLMs (Anthropic, OpenAI), Python (scripts and integrations), automations.

DOM · **Head of Marketing**

Jun 2022 – Jul 2023 · João Pessoa, PB, Brazil

Marketing leadership and implementation of internal technology infrastructure. Creation of an intranet, customer service automations, projects for domestic and international clients. Team training and development.

CVC Viagens · **Marketing Coordinator and Manager**

Mar 2017 – Jun 2021 · PB/PE, Brazil

Structured and led the marketing hub responsible for all stores in Paraíba and the main store in Pernambuco (Shopping Recife). In 2025, invited to return as strategic consultant for the PE-PB hub plan: 68 stores, R\$ 285M (BRL) in operating budget, projected ROI of 11.7x, payback of 3.1 months.

Previous positions: TagZag (Account and Planning Executive, 2021–2022); Classic Viagens (Marketing Manager, 2017–2021); Experimento Intercâmbio (Sales Manager, 2017–2020); Folks Comunicação (Co-founder, 2015–2017); Soda Virtual (Planning Manager, 2013–2015); PB Turismo (Travel Agent, 2011–2013); Societá Viagens (2010–2011); CVC Viagens (Sales Associate, 2008–2010).

Classic Viagens and Experimento Intercâmbio belonged to the same CVC franchise group in Paraíba; I served simultaneously as marketing manager for all three fronts and sales manager for Experimento.

APPLIED ENGINEERING ON PERSONAL PROJECTS

- wagnerlima.cc: personal site in Next.js + MDX on Vercel, fed by an Obsidian vault (second brain) through a deterministic publishing pipeline in Python: three publishers with frontmatter gates, internal link validation, category allowlist, and invisible AI mark checking via a local MCP server. Interactive vault graph generated by a Python script (302 nodes, edges extracted from the notes, private content anonymized by salted hash) and rendered with animation on the site.
- Writing workflows orchestrated by Claude Code skills (reading → polishing → external adversarial audit → publishing), with single canonical docs in the vault and authorship rules that separate AI research from the author's voice. Scheduled tasks and vault linting in Python. Open-source models (Llama, Mistral) run on my own hardware for experimentation.

TECHNICAL STACK

Languages: TypeScript, JavaScript/Node.js, C#/.NET, Python, Go, SQL.

AI and LLMOps: advanced prompt engineering, prompt caching and system prompt compression, governance, observability, cost, and traceability in production, integration with LLM APIs (Anthropic, OpenAI), building MCP (Model Context Protocol) servers, A2A protocol, iterative prompt evaluation pipeline in production, architectures where the LLM interprets and the code calculates, local open-source models.

Agents: driving coding agents (Claude Code) through milestone-based builds with technical and security review checkpoints; orchestrator skills; scheduled tasks; interface contracts between agents with separate territories.

Backend and frontend: NestJS, Next.js, Astro, Prisma, Drizzle, Auth.js, Tailwind, MDX.

Data: PostgreSQL, SQLite, Cloudflare D1, Neon, data science applied to unstructured data.

Infrastructure: Cloudflare Workers, D1, R2, KV, OpenNext, Vercel, Render, Azure, Docker, Linux, Git, CI/CD.

Integrations: Meta WhatsApp Business Cloud API, Telegram, Chatwoot, Resend, Binance API, Microsoft Forms, Todoist via MCP.

Quality: Vitest (Cloudflare plugin), Playwright, Lighthouse, root cause analysis in production incidents.

Differentiators: translating business demands into viable technical architecture; documented handover (runbook, operations manual, handoff) until client autonomy; architecture documentation generated from the code; product decisions guided by real operation.

EDUCATION

Marketing · Estácio de Sá (2020 – 2022)

Social Communication (Advertising) · IESP / UNIESP (2012 – 2016)

LANGUAGES

Portuguese (Native) · English (Fluent)