Abner Evandro Duarte

AI Engineer | Backend Developer

São José do Rio Preto, SP, Brazil | +55 17 99181-6083 | abnerduartw@gmail.com
LinkedIn: linkedin.com/in/abner-evandro-duarte-79352523a | GitHub: github.com/Abner-Duarte
Portfolio: abner-duarte.vercel.app | ZapSync: zapsync-ia.com.br
Open to remote, hybrid or on-site work

SUMMARY

AI engineer who ships LLMs to production. I founded, built and run ZapSync on my own, a multi-tenant SaaS for AI customer service on WhatsApp, with hybrid RAG, structured output, guardrails, evals and model benchmarking, cost and latency observability, and an MCP server. I build backends with Node.js, Python and PostgreSQL, backed by automated tests and CI on GitHub Actions. Before that, almost 3 years in L2/L3 support and ERP implementation in production, with more than 10,000 tickets resolved and a focus on root cause analysis, SQL and logs. Deep knowledge of Brazilian tax systems (NF-e, NFC-e, SEFAZ) and payment terminals (TEF). AWS Certified AI Practitioner, Santander Data with Python and AI Bootcamp, and advanced English (C1).

EXPERIENCE

Founder and AI Engineer | ZapSync

Oct 2025 – Present | São José do Rio Preto, Brazil | zapsync-ia.com.br (public demo at /demo)

Multi-tenant SaaS for AI customer service on WhatsApp that I designed, built and operate on my own: the AI serves customers of restaurants, pizzerias, barbershops and clinics, takes orders, books appointments and hands off to a human when needed.

- Built the message pipeline: HMAC-verified webhook, deduplication, batching of back-to-back messages, per-conversation queue, audio via Whisper, images via a vision model, LLM call, guardrails and database validation.
- Designed an 11-step fallback cascade across OpenAI and Groq (gpt-5.4-mini, gpt-oss-120b and gpt-oss-20b) with strict JSON Schema structured output, so customers never go unanswered.
- Implemented hybrid RAG over the menu in PostgreSQL: keyword search plus embeddings with cosine similarity, incremental re-embedding of changed items only, and keyword-only fallback if the embeddings API goes down.
- Wrote self-correcting guardrails and code-level rules (prices and quantities recomputed from the database, orders only close after a summary), with every trigger logged.
- Built an eval set of 58 conversation turns across 7 categories (12 taken from real production bugs) with a deterministic scorer, plus a per-model benchmark: 97% accuracy, 1.6 s typical response time and US$ 5.45 per 1,000 replies. The benchmark exposed a prompt injection flaw, which I fixed in both the prompt and the code: the security category went from 89% to 100%.
- Implemented LLM observability (cost, tokens, latency, fallback and guardrails per call, via AsyncLocalStorage) shown in a dashboard tab for each business owner.
- Developed an MCP server with 16 tools that lets owners run their business by chatting with an AI assistant, reusing the API's authentication, tenant isolation and audit log.
- Ensured multi-tenant isolation, security (scrypt, signed sessions, rate limiting, LGPD data retention) and 194 automated tests against a real PostgreSQL database on GitHub Actions. Self-service onboarding with Meta Embedded Signup and recurring billing via Mercado Pago. Hosted on Railway and Vercel.

L2/L3 Technical Support Analyst and Implementation Specialist | Consumer Soluções em Tecnologia

Feb 2024 – Sep 2026 | São José do Rio Preto, Brazil

L2 and L3 technical reference for an ERP platform for the food service industry, responsible for keeping restaurants running in production.

- Resolved more than 10,000 support tickets, averaging over 400 per month, from emergency calls to escalated critical cases.
- Diagnosed complex failures through log analysis, direct SQL queries on Firebird databases and tracing of communication between systems, reaching the root cause (RCA).
- Built D3code, an internal AI tool adopted by the support team: it transcribes customer voice messages (Groq Whisper) and extracts data from screenshots (Gemini), running on Cloudflare Workers with Google OAuth restricted to the company domain and per-user usage tracking.
- Reproduced, documented and reported bugs to the development team, validating fixes in customer environments after each release.
- Handled critical integrations: NF-e and NFC-e e-invoicing with SEFAZ, digital certificates, TEF payment terminals and pinpads, and delivery platforms (iFood, Keeta, 99Food).
- Led end-to-end implementations (requirements, installation, tax setup, data migration and go-live), standardized procedures, and trained new team members and customer staff.

Independent IT Technician (MEI)

On-demand services

- Installation and configuration of IT equipment and infrastructure, including TV installations in schools, with formal invoicing.

PROJECTS

Data Pilot (Santander/DIO Bootcamp project): a sales dashboard you can talk to. The LLM picks which Python function to call (tool calling) and Python does the math, so the model never makes up numbers. Tests guarantee the chat and the dashboard show the same value.
Python, Streamlit, Plotly, OpenRouter, pytest · github.com/Abner-Duarte/data-pilot

Credit card fraud detection (Santander/DIO Bootcamp project): 284k transactions with 0.17% fraud. Compared models by precision, recall and PR-AUC, tuned the decision threshold and explained predictions with SHAP; the final XGBoost catches 83 of 98 frauds with 15 false alarms.
Python, scikit-learn, XGBoost, SHAP, Jupyter · github.com/Abner-Duarte/deteccao-fraude-cartao

BeHappier: Android app with multimodal AI (chat, vision and voice) that I designed and built on my own, with about 8,700 lines of TypeScript, a Firebase backend and the API key protected in Cloud Functions.
React Native, Expo, TypeScript, Firebase, OpenAI · github.com/Abner-Duarte/behappier-showcase

QA take-home assignment (Verzel): test plan, Gherkin test cases, manual and exploratory testing, and Playwright automation (UI on Chrome, Firefox, Safari and mobile, plus API tests) running on GitHub Actions. Found 4 bugs, 2 of them high severity.
Playwright, TypeScript, Gherkin, GitHub Actions · github.com/Abner-Duarte/Teste-qa-verzel

ZapSync lead generation: script that searches local businesses by category and city through the Google Places API and builds the outreach contact list; the first run collected 1,500 leads in São José do Rio Preto and Mirassol.
Python, Google Places API, CSV

Quest Contratado: web game I built to gamify my job search and studies, with XP, daily quests, Python (60 lessons), SQL and AWS tracks with spaced repetition, a 112-lesson cloud roadmap, AI mock interviews and an animated 3D hero.
JavaScript, HTML, CSS, three.js, three-vrm

Reinos Eternos: 3D fantasy idle RPG in the browser, with 4 animated classes, combat with armor, dodge and block, talent trees with synergies, item forging, 10 zones and 24 monsters, plus my own simulator to balance progression.
JavaScript, Three.js, WebAudio, Node.js

TT2 Optimizers: toolkit that runs a mobile game's optimizers locally from the account export, with my own engine in Node.js and Python, the game's official CSV tables and headless execution of web pages.
Node.js, Python, jsdom, CSV, JSON

Chatwoot userscript (Consumer): browser extension that listens to Chatwoot's real-time events and alerts agents when a new conversation enters the queue, with no server changes.
JavaScript, Tampermonkey, WebSocket

QA study simulators (Consumer): interactive HTML simulators I built to prepare for an internal QA position, covering Azure DevOps, Selenium, Cypress, Robot Framework, Git, SQL and Postman.
HTML, CSS, JavaScript

Python games: two Pygame Zero games with beginner-friendly commented code, with all art and sound generated by code, plus three terminal scripts (inventory, sales commission and late-payment interest).
Python, Pygame Zero, Pillow · github.com/Abner-Duarte/portfolio-python

Python coding challenge: three business-rule exercises (sales commission from JSON, inventory movements with validation and late-payment interest) using only the standard library.
Python, json, datetime, uuid · github.com/Abner-Duarte/Desafios-Dev

Arduino gate simulator (FATEC, 2026): academic project simulating an automatic gate with Arduino, documented in a blog.
Arduino, electronics

WhatsApp bot with GPT: my first WhatsApp bot connected to a GPT model, which became the foundation for ZapSync.
WhatsApp, OpenAI GPT

Barbearia do Pra (FATEC, team project): full-stack web scheduling system with an admin panel for bookings, barbers, services and opening hours.
Java 17, Spring Boot, SQL Server, JavaScript · github.com/Abner-Duarte/Barbearia-Site

Fatec Fighters (FATEC, team project): 2D two-player fighting game with login, scoring and a leaderboard in PostgreSQL. It evolved from a single-class prototype into separate screens with a DAO layer and parameterized queries.
Java, libGDX, PostgreSQL, JDBC, Gradle · github.com/Abner-Duarte/FatecFighters

SysOs (front-end course final project, 2023): marketing and sales website for a work-order and online-ordering system for small businesses, with pricing plans, sign-in/sign-up and a responsive layout.
HTML, CSS, JavaScript, Bootstrap · github.com/Abner-Duarte/Projeto-Final

TECHNICAL SKILLS

AI and LLMs: OpenAI, Groq, Gemini, Whisper, prompt engineering, structured output (JSON Schema), tool calling, agents, hybrid RAG, embeddings, vector search, guardrails, evals and model benchmarking, LLM observability, prompt injection, MCP (Model Context Protocol)

Machine learning and data: pandas, scikit-learn, XGBoost, SHAP, metrics for imbalanced classes, Streamlit, Plotly, Jupyter

Backend: Node.js, Express, Python, Java, Spring Boot, REST APIs, webhooks, WhatsApp Cloud API, multi-tenant architecture, API authentication and security

Databases: PostgreSQL, SQL Server, Firebird, Firestore, data modeling, production SQL troubleshooting, data migration

Front-end and mobile: JavaScript, TypeScript, HTML, CSS, React Native, Expo

QA and testing: Playwright, Gherkin (BDD), API testing, functional, regression and exploratory testing, node:test, pytest, bug reproduction and reporting, acceptance criteria

Cloud and DevOps: GitHub Actions (CI/CD), Railway, Vercel, Firebase, Cloudflare Workers, Git, AWS (AI Practitioner)

Support and business domain: L2/L3 support, root cause analysis, ERP implementation, go-live, Brazilian e-invoicing (NF-e, NFC-e, SEFAZ), digital certificates, TEF, pinpads, POS, food service, user training

EDUCATION

Associate Degree in Information Technology for Business (Tecnólogo em Informática para Negócios) · FATEC (São Paulo State Technological College), São José do Rio Preto · Completed 2026
Thesis on automatic shutdown mechanisms in AI.

CERTIFICATIONS AND COURSES

- **AWS Certified AI Practitioner (AIF-C01)**, Amazon Web Services
- **Santander Data with Python and AI Bootcamp**, DIO, completed Nov 2026: Python and OOP, testing and best practices, data cleaning, visualization and automation, relational databases, machine learning, NLP, computer vision, generative AI, prompt engineering and AI agents
- **IBM Confluent Bootcamp: Real-Time Data for AI Agents**, DIO
- EF SET English Certificate: 69/100 (C1 Advanced)
- NodeJS: The Complete Guide (incl. MVC, REST APIs, GraphQL), Udemy (Maximilian Schwarzmüller)
- 100 Days of Code: The Complete Python Pro Bootcamp, Udemy (Angela Yu)
- SQL and PostgreSQL: The Complete Developer's Guide, Udemy (Stephen Grider)
- Playwright JS/TS Automation Testing from Scratch & Framework (Cucumber BDD, CI/CD with GitHub Actions), Udemy (Rahul Shetty)
- The AI Engineer Course: Complete AI Engineer Bootcamp, Udemy (365 Careers)

LANGUAGES

Portuguese: native · English: advanced (C1)