CARLOS HENRIQUE COSTA DA SILVA JUNIOR

AI ENGINEER | GENAI, LLMs E ENGENHARIA DE SOFTWARE (PYTHON, FASTAPI, RAG, AGENTES)

Rio de Janeiro, RJ - Remoto | (21) 98883-5047 | chjunior05@gmail.com
linkedin.com/in/carlos-henrique-junior | github.com/Carlos-Henrique-Junior | carlos-henrique-junior.github.io

RESUMO PROFISSIONAL

Engenheiro de software com foco em IA generativa, atuando na construção e integração de aplicações que usam LLMs de ponta a ponta: APIs em Python e FastAPI, orquestração de agentes com LangChain e LangGraph, pipelines de RAG com embeddings e banco vetorial, prompt engineering e avaliação de respostas. Na PGE-RJ desenvolvi o FastAPI Radar, API que centraliza métricas de monitoramento e alimenta dashboards, automações e consultas de negócio, além de pipelines de ETL em Apache Airflow e SQL. Cinco anos de experiência em TI, com suporte N2, SLA e operação de ambientes Linux. Pós-graduações em Inteligência Artificial e em Machine Learning (2026) e bacharelado em Ciência da Computação (2025). Disponibilidade de 30 horas semanais para atuação remota.

COMPETÊNCIAS TÉCNICAS

IA generativa e LLMs: OpenAI (GPT), Anthropic (Claude), Google (Gemini), LangChain, LangGraph, RAG, embeddings, banco vetorial, prompt engineering, function calling, Model Context Protocol (MCP), fine-tuning, Hugging Face

Linguagens e backend: Python, FastAPI, APIs REST, arquitetura de APIs, Pydantic v2, SQLAlchemy 2 (assíncrono), Alembic, autenticação JWT, Django (curso), SQL, Bash, JavaScript

Dados e machine learning: Pandas, NumPy, Apache Airflow, ETL, automação de rotinas, modelagem de dados, análise exploratória, métricas e KPIs, Power BI, Streamlit, Plotly

Banco de dados: PostgreSQL, MySQL, SQL Server, SQLite, modelagem relacional, otimização de consultas SQL

Infraestrutura e entrega: Docker, Docker Compose, deploy, Git/GitHub, CI/CD (GitHub Actions), Linux, AWS (EC2)

Qualidade e observabilidade: pytest, httpx, Ruff, Grafana, Zabbix, documentação técnica

EXPERIÊNCIA PROFISSIONAL

Técnico de Suporte N2 | Montreal Informática
04/2026 - 07/2026 | Rio de Janeiro, RJ

- Sustentei sistemas corporativos em produção, com diagnóstico e resolução de incidentes e cumprimento de SLA.
- Automatizei rotinas de verificação e extração de dados com Python e SQL, reduzindo etapas manuais do atendimento.
- Documentei procedimentos e fluxos de diagnóstico, criando base de conhecimento reutilizada pela equipe.

Estagiário de TI / Data Analytics | Procuradoria Geral do Estado do Rio de Janeiro (PGE-RJ)
03/2024 - 03/2026 | Rio de Janeiro, RJ (híbrido)

- Desenvolvi o FastAPI Radar, API REST em Python e FastAPI que centraliza métricas de monitoramento e alimenta dashboards, automações e consultas de negócio de diferentes áreas.
- Construí pipelines de ETL com Apache Airflow e Python, integrando fontes heterogêneas em modelo de dados analítico usado por relatórios e indicadores de operação.
- Modelei e otimizei consultas SQL em MySQL para relatórios e dashboards, com controle de acesso a informações sensíveis e observância à LGPD.
- Automatizei rotinas operacionais com Python, Bash e SQL, substituindo execuções manuais por scripts versionados no Git e reduzindo etapas repetitivas do time.
- Operei monitoramento com Zabbix e Grafana, apoiando diagnóstico de incidentes e acompanhamento de disponibilidade dos sistemas.

Técnico de Suporte em TI | CREB - Centro de Reumatologia e Ortopedia Botafogo
12/2021 - 12/2023 | Rio de Janeiro, RJ

- Prestei suporte presencial e remoto a mais de 200 usuários em unidades de saúde, cobrindo hardware, software, redes e os sistemas Tasy e GLPI.

Estagiário de TI | CIOB - Centro de Integração Objetivo
01/2020 - 12/2020 | Rio de Janeiro, RJ

- Mantive computadores e a infraestrutura de TI do centro educacional, com instalação e configuração de Windows e Linux.

PROJETOS

ITSM com IA assistiva | Django 6, Python, PostgreSQL, LangGraph, Celery, Docker, CI

- Plataforma de gestão de serviços de TI multi-tenant em desenvolvimento: chamados com fila, prioridade, SLA por política e linha do tempo imutável de atendimento.
- Fundação entregue com isolamento entre organizações coberto por testes, anexos protegidos em rota autenticada, Docker Compose e pipeline de CI no GitHub Actions.
- Assistente de IA sobre o histórico de chamados, em implementação com LangGraph e RAG: classificação, resumo e sugestão de solução com citação das fontes e aprovação humana.

MADR - Acervo digital de livros | Python, FastAPI, PostgreSQL, Jinja2, HTMX, Docker, CI

- Aplicação completa: API REST com autenticação JWT, SQLAlchemy 2 assíncrono e migrações com Alembic, mais interface web com login, painel de métricas, CRUD, busca, filtro e paginação.
- Mais de 90 testes automatizados com pytest e httpx, container Docker com healthcheck e pipeline de CI no GitHub Actions.

Monitoramento de Mercado - COT Report (CFTC) | Python, Pandas, SQL Server, FastAPI, Streamlit, Docker

- Pipeline de ETL que extrai, trata e persiste o relatório público de posicionamento institucional (Commitments of Traders) em base histórica consultável.
- API em FastAPI sobre o data warehouse e dashboard Streamlit publicado, com indicadores de posicionamento de fundos e análise estatística de tendência.
- Consumo dos mesmos dados em Power BI e Excel, reaproveitando o modelo em vez de repetir extração.

Products API - Gestão de inventário com dashboard | Python, FastAPI, PostgreSQL, Streamlit, Docker, CI

- API com autenticação JWT, migrações Alembic, testes com cobertura de branch e lint com Ruff.
- Dashboard em Streamlit e Plotly para acompanhar inventário, com Docker Compose e CI no GitHub Actions.

API de Gerenciamento de Voluntários | Python, FastAPI, PostgreSQL, SQLAlchemy, Alembic, Docker, CI

- API REST para cadastro e controle de voluntários, com validação em Pydantic v2, SQLAlchemy 2 assíncrono, migrações Alembic e configuração por ambiente.
- Testes com pytest, pytest-asyncio, httpx e cobertura, Dockerfile multi-stage, Docker Compose e pipeline de CI no GitHub Actions.

Análise de salários do mercado de dados | Python, Pandas, NumPy, Matplotlib, Seaborn, Plotly, Streamlit

- Análise exploratória de base pública com mais de 130 mil registros, com limpeza, padronização e tradução de variáveis para investigar o efeito de senioridade, modelo de contratação, localização e porte da empresa.
- Dashboard interativo com testes automatizados de execução e de contrato do conjunto de dados.

PRÁTICA EM IA GENERATIVA

- Integração de APIs de LLM em Python (GPT, Claude, Gemini), com function calling, saída estruturada em JSON, controle de tokens e tratamento de erro e de timeout.
- Pipelines de RAG: chunking, geração de embeddings, busca semântica em banco vetorial, re-ranking e resposta com citação da fonte.
- Orquestração de agentes com LangChain e LangGraph e construção de agente com ferramentas próprias expostas via MCP (Model Context Protocol) para consultar dados e executar tarefas.
- Prompt engineering com versionamento de prompts, exemplos few-shot e testes de regressão de respostas para medir qualidade antes de publicar mudança.
- Experimentos de treinamento e ajuste fino (fine-tuning) de modelos open-source do Hugging Face, comparando custo, latência e qualidade com modelos de API.

FORMAÇÃO ACADÊMICA

- Pós-graduação em Inteligência Artificial - Faculdade Líbano, 2026
- Pós-graduação em Machine Learning - Faculdade Líbano, 2026
- Bacharelado em Ciência da Computação - Universidade Estácio de Sá, conclusão em 12/2025
- Curso Técnico em Informática - CIOB - Centro de Integração Objetivo, 2018

CERTIFICAÇÕES E CURSOS

- Desenvolvimento Web com Python e Django - DIO.ME
- Imersão IA para Devs - PyCodeBR
- Docker e Fundamentos de Linux - DIO.ME e Udemy
- Introdução à Engenharia de Dados com Python - DIO.ME
- Em preparação: AWS Certified Cloud Practitioner

IDIOMAS E DISPONIBILIDADE

Português nativo | Inglês intermediário (B1) - leitura técnica e documentação | Espanhol básico.

Disponibilidade imediata para trabalho remoto, 30 horas semanais.