E-mail

contato@praxis-edu.app

Localização

Rio de Janeiro, Brasil

GitHub

github.com/
MatheusOliveiraSilva

LinkedIn

linkedin.com/in/
matheus-silva

Idiomas

- Português (nativo)
- Inglês (profissional)

Formação

M.Sc., IA Generativa

PUC-Rio

B.Sc., Ciência da Computação

Universidade Federal do Rio de Janeiro (UFRJ)

Matheus Oliveira Silva

ENGENHEIRO DE SOFTWARE, AI PLATFORM
AI AGENTS · SISTEMAS DE LLM · DEVELOPER TOOLS

Resumo Profissional

Engenheiro de Software com visão de produto e mais de 7 anos de experiência em engenharia de backend, IA aplicada, pesquisa, plataformas de dados e LLMOps. Especializado em plataformas de IA em produção e sistemas agênticos, incluindo execução autônoma, memória de longo prazo e reconstrução de contexto, orquestração multiagente, ferramentas para desenvolvedores, avaliação e observabilidade para agentes confiáveis de longa duração.

Experiência Profissional

ENGENHEIRO DE SOFTWARE II, AI PLATFORM & LLMOPS

mar 2026 - atual

Housecall Pro
Remoto

- Entreguei todo o sistema de evals da plataforma de agentes e liderei o processo de evals, que funciona dentro do AI Manager, o mesmo software em que os agentes da plataforma são gerenciados.
- Criei o plugin MCP que permite aos usuários se conectar via MCP para definir itens de dataset, rodar evals e melhorar agentes e prompts com base nos resultados, cobrindo todo o ciclo de desenvolvimento de agentes.
- Entreguei a camada de memória conversacional de longo prazo: transcrições duráveis em PostgreSQL para eventos de modelo, ferramentas, subagentes e citações; reconstrução do contexto do LLM após a expiração de checkpoints do LangGraph; e APIs de restauração de chat com paginação por cursor.
- Liderei o desenvolvimento de agentes internos que automatizam o SDLC e o debugging e a triagem de issues a partir de alertas do Datadog (LangGraph, Deep Agents, LangSmith, análise de código em sandbox e runtimes isolados em Kubernetes), aumentando a eficiência operacional.

ENGENHEIRO DE SOFTWARE, VOICE AI & SISTEMAS MULTIAGENTE

set 2025 - mar 2026

Simbie AI
Nova York, Estados Unidos · Remoto · Startup apoiada pela Y Combinator

- Rearquitetei um MVP de voice AI, antes um wrapper simples do Retell AI em torno de um prompt monolítico, transformando-o em um backend de orquestração multiagente em TypeScript com agentes especializados, roteamento explícito e handoffs.
- Construí a integração em tempo real via WebSocket com a camada de telefonia do Retell AI, coordenando a execução dos agentes em chamadas telefônicas ao vivo.
- Implementei observabilidade baseada em Grafana para sessões de agentes de voz, apoiando o monitoramento e o debugging em produção do ciclo de vida das chamadas, da latência e de falhas.

Competências Técnicas

IA e Sistemas de Agentes

- AI agents
- LangGraph
- Deep Agents
- MCP
- Orquestração multiagente
- Tool use
- Context engineering
- Memória de longo prazo
- RAG, GraphRAG
- Avaliação de LLMs

Linguagens e Backend

- Python, TypeScript, Go
- FastAPI
- WebSockets
- APIs REST
- PostgreSQL, Redis

Plataforma e Observabilidade

- Kubernetes, Docker
- AWS
- GitLab CI/CD
- HashiCorp Vault
- Datadog, LangSmith, Grafana

Publicações

"Automated Evaluation of Database Conversational Agents."

Silva, M. et al. · WEBIST 2025, primeiro autor, publicado.

"A Database Dialogue Interface Leveraging LLMs."

Silva, M. et al. · ER 2025, primeiro autor, aceito.

"The Quest for an LLM-based Natural Language Interface for an Industrial Database."

Silva, M. et al. · PAIS 2025, coautor, aceito.

Experiência Profissional (cont.)

ENGENHEIRO DE IA E DADOS

mar 2024 - set 2025

BTG Pactual

Rio de Janeiro, Brasil · Híbrido

- Arquitetei um runtime de agente Text-to-SQL em produção, com definições declarativas de agentes no PostgreSQL compiladas em grafos de execução do LangGraph e servidas via FastAPI, com interfaces de streaming e de voz.
- Com o agente, os tickets de suporte caíram de cerca de 20 por semana para zero, aumentando a eficiência operacional.
- Construí uma camada de orquestração MCP conectando ferramentas internas de recuperação e execução; implementei estratégias de RAG e GraphRAG para consultas que respeitam as regras de negócio sobre datasets analíticos complexos.
- Entreguei um pipeline de avaliação offline de LLMs com guardrails de release que bloqueavam versões de agentes com regressão em qualidade de resposta, latência ou custo.
- Projetei e operei uma plataforma de enriquecimento de dados orientada a eventos com AWS Step Functions e Lambda, para workflows de alto throughput e baixa latência que alimentavam a camada de agentes.

ENGENHEIRO DE SOFTWARE DE PESQUISA

abr 2023 - mar 2024

Instituto Tecgraf · PUC-Rio

Rio de Janeiro, Brasil · Remoto

- Transformei pesquisa em Text-to-SQL em um agente conversacional implantado internamente na Petrobras para milhares de engenheiros, definindo a arquitetura e liderando a implementação.
- Primeiro autor de pesquisa sobre avaliação automatizada de agentes conversacionais para bancos de dados e coautor de trabalho sobre Text-to-SQL industrial.

ENGENHEIRO DE SOFTWARE, BACKEND & IA APLICADA

mar 2020 - abr 2023

cogg.ai

Rio de Janeiro, Brasil · Remoto · Promovido de Estagiário de Engenharia de Software (mar 2019 - mar 2020)

- Backend em Python e AWS (serviços, APIs, sistemas de recomendação e aplicações web) para produtos de clientes, além de chatbots com LLM usando as APIs da OpenAI já no início da adoção do ChatGPT.

Liderança e Ensino

INSTRUTOR E EMBAIXADOR DE AI ENGINEERING

fev 2025 - atual

Borderless Coding

- Instrutor convidado recorrente, ensinando AI engineering e desenvolvimento de software agêntico, incluindo agentes de LLM, workflows de programação assistida por IA, agent skills, plugins, hooks e ferramentas para desenvolvedores.
- Conduzo mock interviews de AI System Design e faço mentoria de profissionais brasileiros de tecnologia que buscam carreiras globais.

Projetos em Destaque

CURSOR SDK CODING AGENTS

2026 - atual

Projetos independentes

- Desenvolvo coding agents de longa duração para exploração de repositórios, implementação com escopo definido, validação automatizada e workflows de entrega baseados em Git, usando o Cursor SDK, skills, plugins e hooks.