

JAIR PEREIRA DA SILVA JÚNIOR

AI Engineer | LLMs | AI Agents | RAG | Python | FastAPI

E-mail: juniorsilva@msn.com - WhatsApp: (31) 997838681

LinkedIn: [linkedin.com/in/jairengdados](https://www.linkedin.com/in/jairengdados) | GitHub: github.com/Jair-pc | Portfólio: <http://www.jrdados.com.br/>

OBJETIVO PROFISSIONAL

Atuar como AI Engineer no desenvolvimento de aplicações baseadas em IA Generativa, LLMs, agentes inteligentes e sistemas RAG, construindo soluções escaláveis para ambientes de produção e aplicando minha experiência em Engenharia de Dados, arquitetura de software e integração de sistemas.

RESUMO DAS QUALIFICAÇÕES

AI Engineer e Data Engineer com experiência no desenvolvimento de aplicações baseadas em LLMs, agentes inteligentes, sistemas RAG e integrações com APIs. Desenvolve aplicações de IA para produção utilizando Python, FastAPI, PostgreSQL, Docker, OpenAI, Anthropic, Gemini, LangChain e MCP (Model Context Protocol), aplicando Prompt Engineering, recuperação de contexto, observabilidade e avaliação de modelos em aplicações de IA para produção. Também possui sólida experiência em Engenharia de Dados, arquitetura de software, automação e desenvolvimento de produtos SaaS orientados por dados. Atualmente cursa Pós-graduação em IA Generativa e Aplicações com LLMs pela PUC Minas.

PRINCIPAIS COMPETÊNCIAS

- Generative AI
- AI Engineering
- AI Agents
- LLMs
- RAG
- Prompt Engineering
- FastAPI
- Python
- PostgreSQL
- Docker
- RESTful APIs
- Engenharia de Dados

EXPERIÊNCIAS PROFISSIONAIS

Desenvolvedor de IA e Engenharia de Dados | Autônomo - 10/2022 – Atual.

- ✓ **Desenvolvimento de aplicações de IA e plataformas SaaS:** Desenvolvimento de aplicações de IA utilizando LLMs, agentes inteligentes, RAG e integrações com APIs para ambientes de produção.
- ✓ Desenvolvimento de plataformas SaaS utilizando Python, FastAPI, PostgreSQL, Docker e arquitetura modular, integrando IA Generativa para análise, automação e tomada de decisão.

- ✓ Implementação de sistemas RAG com recuperação de contexto, Prompt Engineering, avaliação de respostas, observabilidade, monitoramento de latência e otimização da qualidade das respostas.
- ✓ Construção de pipelines de Engenharia de Dados (ETL), integração de múltiplas fontes de dados e processamento analítico para suporte a aplicações de IA.
- ✓ Desenvolvimento da plataforma +Leite AI Platform para análise técnica, geração de recomendações inteligentes e apoio à tomada de decisão utilizando IA Generativa.
- ✓ Desenvolvimento de agentes inteligentes integrados ao WhatsApp e aplicações web utilizando APIs REST, PostgreSQL e modelos de linguagem.
- ✓ Participação em projetos de renovação tecnológica para os bancos Santander e Itaú Unibanco em parceria com a Aris Tecnologia.

FORMAÇÃO ACADÊMICA

- ✓ **2026** – Pós Graduação - IA Generativa e Aplicações com LLMs – **Cursando - PUC Minas.**
- ✓ **2022** – Pós Graduação - Engenharia de Dados – **PUC Minas.**
- ✓ **2018** – Graduado em Engenharia Elétrica - **UNI BH.**
- ✓ **2010** – Graduado em Publicidade e Propaganda – **Estácio de Sá.**
- ✓ **2015** - Técnico em Eletroeletrônica – **UNA.**

CURSOS E CERTIFICAÇÕES

- ✓ **Idiomas – Inglês** - Gabi Modesto - **Cursando** - 2026.
- ✓ **Arruda data consulting** – Snowflake e dbt - 48h - **Cursando** - 2026.
- ✓ **Jornada de Dados** – Automações com n8n - 50h - 2025.
- ✓ **DSA** - Engenheiro de Dados - **Cursando** - 424h - 2025.
- ✓ **Indicium** - Engenharia de Analytics – 60h - 2025.
- ✓ **Arruda data consulting** - Apache Hop e Metabase - 48h - 2025.
- ✓ **Uni1500** - Inteligência artificial e Análise de Dados utilizando LLM - 160h - 2024.
- ✓ **CoderHouse** - Data Analytics - 48h - 2024.
- ✓ **Xp Educação** - Bootcamp AWS Cloud Computing - 150h - 2022.
- ✓ **Xp Educação** - Bootcamp Azure Cloud Computing - 150h - 2022.
- ✓ **Xp Educação** - Bootcamp Engenheiro de Dados - 150h.150h - 2022.

STACK TÉCNICA:

Linguagens e Desenvolvimento

- ✓ Python (Pandas, NumPy) • SQL • HTML5 • CSS3 • JavaScript

AI Engineering

- ✓ **LLMs:** OpenAI • Anthropic • Gemini • GLM
- ✓ **Frameworks:** LangChain • MCP • OpenClaw
- ✓ **Arquitetura:** AI Agents • RAG • Embeddings • Vector Search
- ✓ **Engenharia:** Prompt Engineering • Context Engineering

✓ **LLMs Avançado:** Function Calling • Structured Outputs • LLM Evaluation • Hermes

Backend

✓ FastAPI • RESTful APIs • JWT

Banco de Dados

✓ PostgreSQL • MySQL • SQL Server • MongoDB • Redis.

BI e Visualização

✓ Power BI (Modelagem DAX, Relatórios) • Streamlit • Metabase • Dashboards web.

Cloud, DevOps e Ferramentas

✓ AWS • Azure • Docker • Git • GitHub • GitHub Actions • Linux • JSON • YAML • Kubernetes (básico) • Terraform (básico).

Automação

✓ Automação de processos com n8n.

PROJETOS EM DESTAQUE:

+Leite AI Platform (Micro SaaS)

Objetivo: Desenvolvimento da plataforma +Leite AI Platform, utilizada por consultores para análise da qualidade do leite, integrando IA, dashboards analíticos e processamento de dados em tempo real.

Tecnologias: Python, FastAPI, PostgreSQL, Docker, OpenAI, Anthropic, Gemini, RAG, LangChain, Prompt Engineering, HTML, CSS, JavaScript.

Principais funcionalidades:

- Agente de IA especializado em qualidade do leite.
- Recuperação de contexto utilizando RAG.
- Integração com APIs de LLMs.
- Dashboards analíticos em tempo real.
- Geração automática de recomendações técnicas.
- Arquitetura preparada para produção.
- Desenvolvido em FastAPI e PostgreSQL com arquitetura preparada para produção.

Resultado: Plataforma utilizada em ambiente real, reduzindo o tempo de análise dos consultores e centralizando indicadores, documentos técnicos e recomendações inteligentes em um único sistema.

Lumax WhatsApp Agent

Objetivo: Desenvolver um agente conversacional integrado ao WhatsApp capaz de responder clientes utilizando contexto, histórico de conversas e modelos de linguagem.

Tecnologias: Python, FastAPI, Claude API, OpenAI API, PostgreSQL, Docker, REST APIs, Prompt Engineering.

Principais funcionalidades:

- Atendimento automatizado via WhatsApp.

- Contexto persistente das conversas.
- Integração com APIs externas.
- Respostas personalizadas utilizando LLMs.
- Monitoramento e melhoria contínua dos prompts.

Resultado: Agente conversacional com foco em produção, automatizando atendimentos e fornecendo respostas contextualizadas utilizando LLMs.

Assistente Pessoal de IA (OpenClaw + Hermes)

Objetivo: Construir uma plataforma de agentes inteligentes para automatizar tarefas do dia a dia e estudar arquiteturas modernas de AI Engineering.

Tecnologias: OpenClaw, Hermes, MCP (Model Context Protocol), Python, Docker, FastAPI, LangChain, PostgreSQL.

Principais funcionalidades:

- Arquitetura multiagentes.
- Memória persistente.
- Planejamento e execução de tarefas.
- Integração com ferramentas externas.
- Automação de processos pessoais.
- Pesquisa em Context Engineering, planejamento de tarefas e arquiteturas multiagentes.

Resultado: Projeto utilizado para pesquisa e experimentação de arquiteturas modernas de AI Engineering, incluindo agentes inteligentes, MCP, memória persistente, planejamento de tarefas e integração com ferramentas.

HABILIDADES COMPLEMENTARES:

- Arquitetura de aplicações baseadas em LLMs.
- Desenvolvimento de agentes inteligentes.
- Prompt Engineering e Context Engineering.
- Integração de APIs e microsserviços.
- Versionamento com Git e GitHub.
- Documentação técnica.
- Resolução de problemas complexos.
- Pensamento analítico.
- Colaboração em equipes multidisciplinares
- Aprendizado contínuo e rápida adaptação a novas tecnologias.