



Case: API de Chat com Agente de IA

Caro candidato, agradecemos o seu interesse em participar do nosso processo seletivo e por dedicar seu tempo para fazer parte do time de #dreamers! ❤️

Este case tem como objetivo a integração de uma API web com um Agente de IA capaz de utilizar ferramentas externas para responder a consultas. O projeto deve ser implementado utilizando FastAPI para a construção da API e o Strands Agents SDK para a orquestração do Agente de IA. Sinta-se à vontade para estruturar a solução da forma que considerar mais adequada, aplicando boas práticas e demonstrando conhecimentos além do básico, isso será levado em consideração durante a avaliação.

Boa sorte!

Desafio

Desenvolver uma API de Chat simples que se conecta a um Agente de IA. O Agente deve ser configurado para utilizar uma tool de Cálculo Matemático para resolver operações. O projeto deve ser configurado para ser executado localmente com Ollama como LLM (com o modelo base que achar mais adequado).

Requisitos Técnicos

O candidato deverá entregar um projeto funcional que atenda aos seguintes requisitos:

1. Configuração do Ambiente

1.1. Variáveis de Ambiente: O projeto deve utilizar um arquivo .env ou variáveis de ambiente para toda a configuração do modelo de linguagem (LLM) e do Agente.

1.2. Dependências: O projeto deve incluir um arquivo requirements.txt ou pyproject.toml listando todas as dependências necessárias (FastAPI, Strands Agents, etc.).

2. Implementação da API (FastAPI)

2.1. Endpoint de Chat: Crie um endpoint POST /chat que aceite uma mensagem de texto do usuário e retorne a resposta do Agente de IA.

- O payload de entrada deve ser um JSON simples com o campo message.
- A resposta deve ser um JSON com o campo response.

2.2. Servidor: A aplicação deve ser executável via unicorn ou um comando similar, conforme as melhores práticas do FastAPI.

3. Implementação do Agente (Strands Agents)

3.1. Agente Básico: Implemente um Agente de IA utilizando o Strands Agents SDK.

3.2. Tool de Cálculo: O Agente deve ser configurado com uma Tool de Cálculo Matemático.

- O Agente deve ser capaz de identificar quando a pergunta do usuário requer um cálculo e utilizar a Tool para obter a resposta.

- Teste de Uso da Tool: O Agente deve responder corretamente a perguntas como: "Quanto é 1234 * 5678?" ou "Qual a raiz quadrada de 144?".

3.3. Conversação: O Agente deve ser capaz de responder a perguntas de conhecimento geral que não exijam a Tool.

Entrega

O candidato deverá entregar um repositório Git contendo:

1. Código-Fonte: Todos os arquivos Python e de configuração (main.py, requirements.txt, etc.).
2. Instruções de Execução: Um arquivo README.md detalhado com os passos para configurar e executar o projeto localmente, incluindo a instalação do Ollama (ou modelo local similar) e a configuração das variáveis de ambiente.
3. Versionamento: .gitignore mapeando arquivos e diretórios que não devem ser versionados.

Obs: Mesmo que não consiga realizar todos os passos, envie para avaliação.

Referências de Documentação

Para a realização deste teste, o candidato deve consultar as seguintes documentações:

1. FastAPI - Documentação Oficial
<https://fastapi.tiangolo.com/>
2. Strands Agents SDK - Documentação
<https://strandsagents.com/latest/documentation/docs/>
3. Ollama

<https://ollama.com/>

4. Python-dotenv - Gerenciamento de Variáveis de Ambiente

<https://pypi.org/project/python-dotenv/>