

Uso de Agentes para Geração de Relatório Personalizado

Descrição

Construir uma solução baseada em Inteligência Artificial Generativa que consiga consultar dados e notícias para gerar relatórios de forma automatizada, trazendo métricas relevantes e as respectivas explicações que ajudem a explicar o cenário atual que as métricas visam medir.

Pré-requisitos

- Conhecimento em python (listas, tuplas, dicionários, laços de repetição, API e tópicos relacionados);
- Conhecimentos sobre IA Generativa (LangChain, LangGraph, Agents, bancos vetoriais, RAG e técnicas relacionadas)
- Manipulação de LLMs

Atividade Prática

Contexto

A Indicium HealthCare Inc. está interessada em criar uma solução baseada em dados que possa ajudar profissionais da área da saúde a ter um entendimento em tempo real sobre a severidade e o avanço de surtos de doenças. Para tanto, irá realizar uma Prova de Conceito (Proof of Concept - PoC) para avaliar a viabilidade de tal solução e você foi contratado para ser o analista responsável por desenvolver a solução.

Para a PoC escolheu-se criar uma solução que forneça métricas através de um relatório gerado por um **agente** que consiga consultar o banco de dados e trazer os valores das métricas, além disso usando outras ferramentas, este agente deve consultar notícias sobre Síndrome Respiratória Aguda Grave (SRAG) em tempo real para embasar as métricas apresentadas e fornecer **explicações/comentários** sobre o cenário relatado. Será utilizado o conjunto real sobre internações por SRAG disponível no Open DATASUS.

As métricas que devem constar no relatório são:

- taxa de aumento de casos
- taxa de mortalidade
- taxa de ocupação de UTI
- taxa de vacinação da população

Além das métricas citadas o relatório deve possuir dois gráficos, um representando o número diário de casos dos últimos 30 dias e outro mostrando o número mensal de casos dos últimos 12 meses.

Descrição dos dados

Os [dados](#), o [dicionário de dados](#) e demais informações podem ser obtidos no site do [Open DATASUS](#). Os dados estão num arquivo no formato CSV que contém aproximadamente 100 colunas e 165.000 linhas. Por se tratarem de dados reais temos muitos problemas de preenchimento incorreto e dados ausentes. Por conta disso, você deve selecionar as colunas que achar pertinentes e aplicar os tratamentos que achar necessários.

Avaliação

A nota atribuída considerando:

- Escolha da arquitetura
- Governança e Transparência (mecanismos de auditoria e registro de decisões dos agentes)
- Guardrails
- Uso de Tools
- Tratamento de Dados Sensíveis
- Clean Code

Entrega Esperada

Link para um repositório do GitHub contendo a solução desenvolvida. O repositório deve estar público e todas as explicações e documentações do projeto devem ser colocadas no README do repositório e/ou relatórios pertencentes ao repositório.

O repositório também deve incluir um PDF com o diagrama conceitual que ilustre a arquitetura da solução, mostrando o Agente Principal (Orquestrador), as Ferramentas (Tools) que ele utilizará, as interações com o LLM, banco de dados e fontes de notícias.

Entregue o que conseguir fazer mesmo que a entrega não esteja completa. Lembre-se feito é melhor que perfeito!

Você tem **7 dias** para realizar o desafio.