

Módulo 4

O que é a IA generativa?

Inteligência Artificial

É o campo da ciência da computação que busca criar computadores inteligentes que podem replicar ou exceder a inteligência humana

Machine Learning

É um subconjunto da IA que permite que os computadores aprendam com os dados existentes e sejam aprimorados com esses dados a fim de tomar decisões ou fazer previsões.

Aprendizado Profundo

É uma técnica de machine learning no qual camadas de redes neurais são usadas para processar dados e tomar decisões

O que é IA generativa

Crie um conteúdo escrito, visual e auditivo considerando os prompts ou dados existentes

Provisionar um recurso do OpenAI do azure

Implantar um modelo no estúdio do Open AI do azure para usá-lo

- Solicite acesso ao serviço OpenAI do azure
- crie um recurso do openAI do azure no portal do azure

Como implantar modelos de IA generativa

Implantar um modelo no estúdio do openAI do azure para usá-lo:

- Você pode implantar uma ou mais instâncias de cada modelo disponível
- O número de implantações depende da sua conta, que você pode ver no portal
- Como alternativa, use a CLI do azure

Desenvolver aplicativos com o serviço Open AI do azure

Como integrar o OpenAI do Azure ao seu aplicativo

Os aplicativos enviam os prompts aos modelos implantados.

As respostas são conclusões.

- Preenchimento - o modelo usa um prompt de entrada e gera um ou mais preenchimentos previstos

- Inserções - o modelo usa a entrada e retorna uma representação de vetor dessa entrada
- ChatCompletion - o modelo recebe entrada na forma de uma conversa de chat (em que as funções são especificadas com a mensagem enviada) e a próxima conclusão do chat é gerada.

ChatCompletion será o ponto de extremidade em que nos concentramos para este curso

Use a Conclusão e Inserções com modelos baseados em GPT-3

Use ChatCompletion com o GPT-35-Turbo e modelos posteriores

Como Usar os SDKs do OpenAI do azure

SDKs específicos de linguagem estão disponíveis para uso em seus aplicativos, tanto em C# quanto em Python.

A estrutura do código segue um padrão semelhante para ambas as linguagens.

Aplicar a engenharia de prompt com o serviço OpenAI do Azure

O que é a engenharia de prompt?

Construção de prompts para:

- Maximizar a relevância e a precisão dos preenchimentos
- especificar a formatação e o estilo dos preenchimentos
- fornecer contexto de conversa
- reduzir o viés e a aumentar a imparcialidade

Conteúdo primário, de suporte e de base

Pistas

- Os indicadores fornecem um ponto de partida no qual os preenchimentos se baseiam
- Use indicadores para definir o escopo e o formato da saída
- Sugestões são particularmente úteis ao gerar código, Por exemplo:
- Forneça a primeira palavra-chave de uma instrução SQL Select
- defina um nome e parâmetros para uma função do Python

Como usar uma mensagem do sistema

Quando a API de ChatCompletions é usada, uma mensagem do sistema define o contexto que pode afetar os preenchimentos.

Cadeia de pensamento

Peça um modelo para dividir a resposta e explicar o raciocínio

Qual esporte é o mais fácil de aprender, porém, mais difícil de dominar? Dê uma abordagem passo a passo dos seus pensamentos, terminando com a resposta

Implementar a Geração Aumentada de Recuperação(RAG) com o Serviço OpenAI do Azure

Como o OpenAI do Azure pode usar seus dados

Configurar fontes de dados

- Usar fonte de dados existente, como um recurso de pesquisa do Azure
- Use o estúdio do OpenAI do Azure para criar fonte de dados, se você ainda não tiver uma
- Ao criar a fonte de dados, é possível usar dados que já estão em sua conta, como armazenamento de blobs

Use o modelo do OpenAI do Azure, que agora usa seus dados para fundamentação

- Converse com modelos de IA como normal
- Se a fonte de dados tiver informações relevantes sobre o prompt, ela usará esses dados
- Você pode especificar se o modelo de IA está limitado apenas à sua fonte de dados

Conecte-se à sua fonte de dados

- Adicione sua fonte de dados ao playground de chat, em configuração do assistente
- Use uma fonte de dados existente ou use esse assistente para criar uma nova
- Uma vez conectado, uma nova sessão de chat será iniciada. Converse normalmente e veja como o modelo de IA faz referência a esses dados.

Uso de API rest do OpenAI do Azure

- Com cada chamada, você precisa especificar os valores da fonte de dados, juntamente com a matriz de mensagens e quaisquer outros parâmetros
- A autenticação na definição da fonte de dados é para seu recurso de pesquisa, não para seu do OpenAI do Azure.

Como usar os SDKs do OpenAI do Azure

SDKs específicos de linguagem estão disponíveis para uso em seus aplicativos, tanto em C# quanto em Python

A estrutura do código segue um padrão semelhante para ambas as linguagens

As fonte de dados compatíveis atualmente são:

- Azure AI search
- Azure cosmos DB para Mongo DB vCore

- Além de outros em pré-visualização, que em breve serão lançados em GA