

O que é a IA Generativa

Fundamentos de IA do Microsoft Azure

Agenda

- ❑ Conceitos básicos de IA gerativa
- ❑ Conceitos básicos do Serviço OpenAI do Azure
- ❑ Explore a IA Gerativa Responsável

Objetivos de Aprendizado

- ❑ Descreva a IA generativa.
- ❑ Descreva os recursos de grandes modelos de linguagem.
- ❑ Entenda como usar o Azure OpenAI para criar soluções generativas de IA.

Conceitos básicos de IA Generativa

O que é IA generativa?

IA: imita o comportamento humano usando aprendizado de máquina para interagir com o ambiente e executar tarefas sem instruções explícitas sobre o que gerar.

O que é IA generativa?

IA generativa: cria conteúdo original, como IA gerativa que foi incorporada a aplicativos de chat. Os aplicativos de IA gerativa usam entrada em linguagem natural e retornam respostas apropriadas em uma variedade de formatos:

O que é IA generativa?



Geração de
linguagem natural



Geração de código



Geração de
imagem

Modelos de linguagem grandes

Os aplicativos de IA gerativa são alimentados por **LLMs (modelos de linguagem grandes)**, que são um tipo especializado de modelo de machine learning que você pode usar para executar tarefas de **PLN (processamento de linguagem natural)**, incluindo:

Modelos de linguagem grandes

- ☐ Determinar sentimento ou classificar de outra forma o texto em idioma natural.
- ☐ Resumir um texto.
- ☐ Comparar várias fontes de texto quanto à similaridade semântica.
- ☐ Geração de nova linguagem natural.

Modelos de linguagem grandes - transformador

A arquitetura do modelo do transformador consiste em dois componentes principais, ou blocos.

Modelos de linguagem grandes- transformador

- ❑ Um bloco *codificador* que cria representações semânticas do vocabulário de treinamento.
- ❑ Um bloco *decodificador* que gera novas sequências de linguagem.

Modelos de linguagem grandes- transformador

- ❑ O texto é *tokenizado* para que cada palavra ou frase seja representada por um token numérico exclusivo.
- ❑ *Inserções* (valores de vetor com várias dimensões) são atribuídas aos tokens

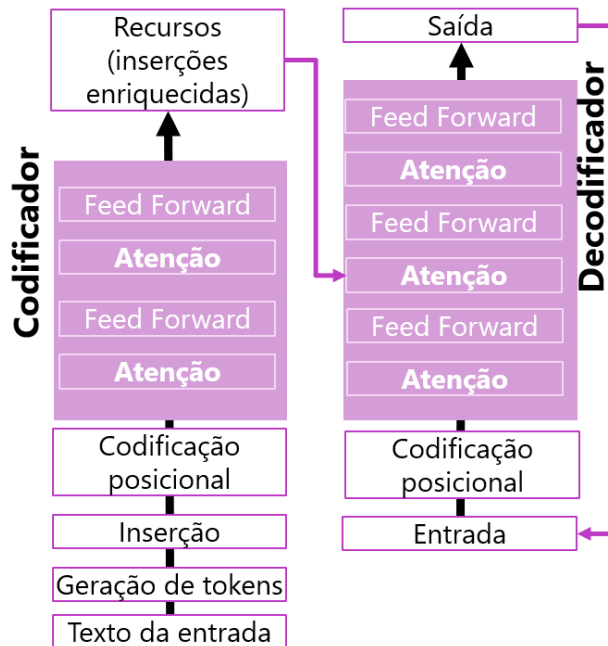
Modelos de linguagem grandes- transformador

- ❑ As camadas de *atenção* examinam cada token por vez e determinam valores incorporados que refletem os relacionamentos semânticos entre os tokens.

Modelos de linguagem grandes- transformador

- ❑ No decodificador, essas relações são usadas para prever a sequência mais provável de tokens.

Modelos de linguagem grandes- transformador



Modelos de linguagem grandes - tokenização

Etapa um: *tokenização*

- ❑ A primeira etapa no treinamento de um modelo de transformador é decompor o texto de treinamento em *tokens*.

Modelos de linguagem grandes – tokenização

Frase de exemplo: *Eu ouvi um cachorro latir alto para um gato.*

"Eu"=1

"ouvi"=2

"um"=3

"cachorro"=4

"latir"=5

"alto"=6

"para"=7

"gato"=8

Grandes modelos de linguagem – tokenização

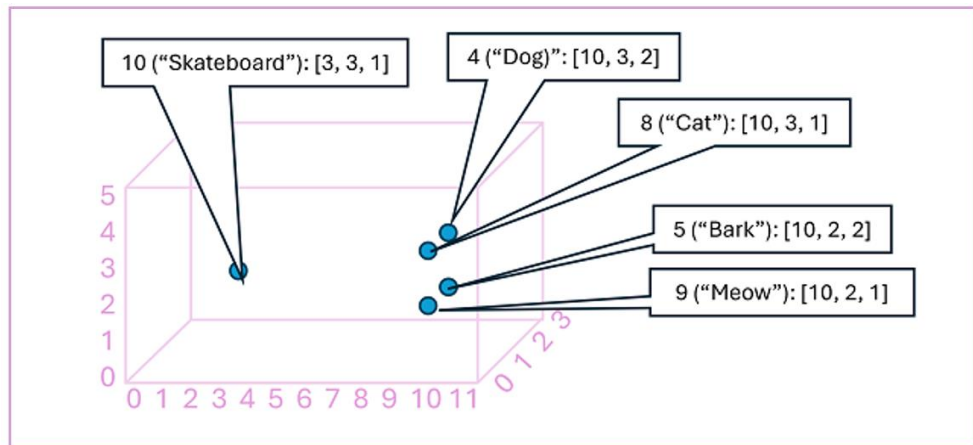
- ❑ A frase agora é representada com os tokens:
[1 2 3 4 5 6 7 3 8].
- ❑ Observe que “um” é tokenizado como 3 apenas uma vez.
- ❑ Da mesma forma, a frase “Eu ouvi um gato” poderia ser representada com as fichas **[1 2 3 8]**.

Modelos de linguagem grandes – inserções

Etapa dois: **inserções**

- ❑ As relações entre tokens são capturadas como vetores, conhecidos como inserções.

Modelos de linguagem grandes – inserções



Token	Word	Inserção
10	Skateboard	[3, 3, 1]
4	Cachorro	[10,3,2]
8	Gato	[10,3,1]
5	Latir	[10,2,2]
9	Miau	[10,2,1]

Fonte: Microsoft

Modelos de linguagem grandes – atenção

Terceiro passo: **atenção**

- ❑ Capture a força das relações entre tokens usando a técnica de atenção.

Modelos de linguagem grandes – atenção

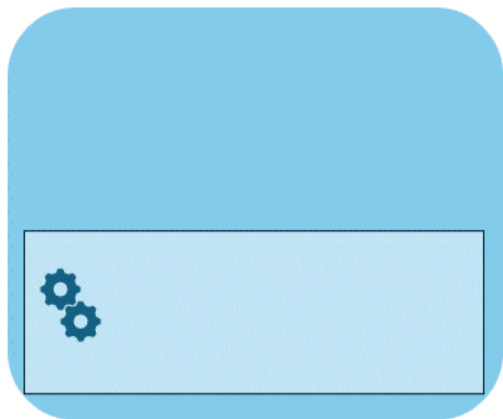
Exemplo:

- ❑ Meta: prever o token após "**cachorro**".
- ❑ Represente "**Ouvi um cachorro**" como vetores.
- ❑ Atribua mais peso a "**ouvi**" e "**cachorro**".

Modelos de linguagem grandes – atenção

- ❑ Vários tokens possíveis podem vir depois de cachorro.
- ❑ O token mais provável é adicionado à sequência, nesse caso , "**latir**".

Modelos de linguagem grandes – atenção



Fonte: Microsoft

Copilotos

- ❑ Os copilotos são frequentemente integrados a outros aplicativos e fornecem uma maneira para os usuários obterem ajuda com tarefas comuns a partir de um modelo generativo de IA.

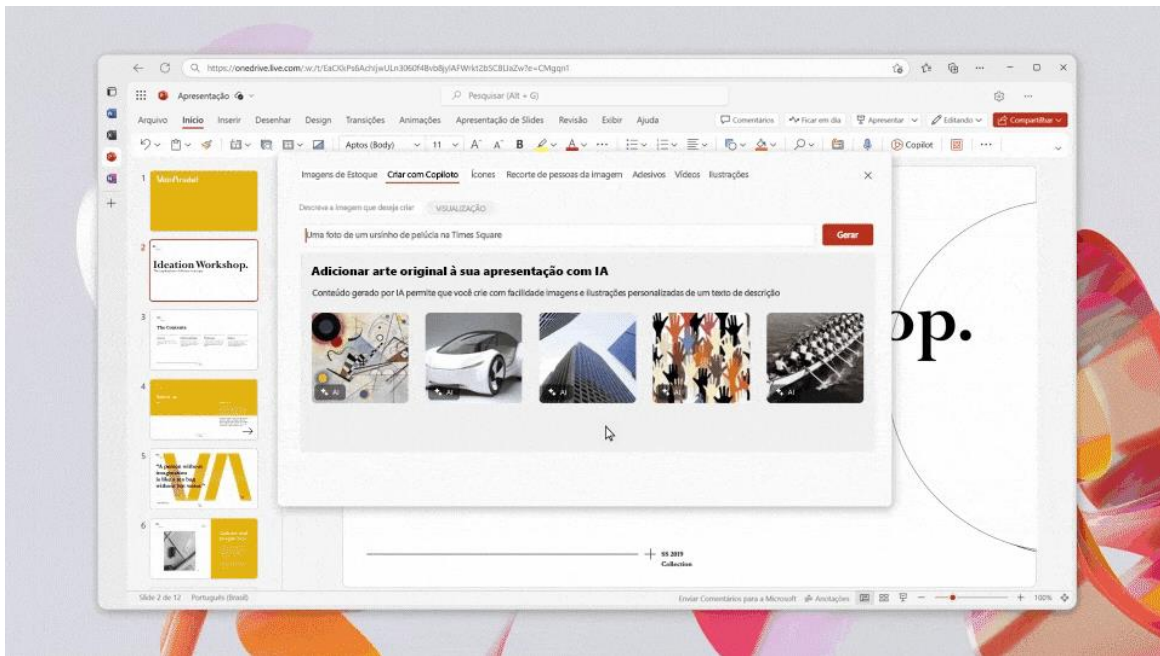
Copilotos

- ❑ Os desenvolvedores podem criar copilotos que enviam prompts para grandes modelos de linguagem e geram conteúdo para uso em aplicativos.

Copilotos

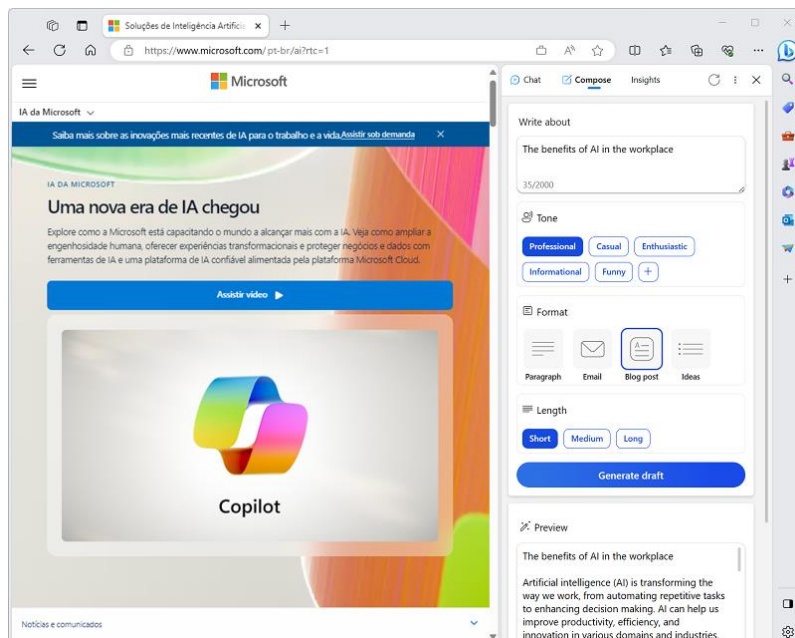
- ❑ Os usuários empresariais podem usar copilotos para aumentar sua produtividade e criatividade com conteúdo gerado por IA.

Copilots



O que são copilotos?

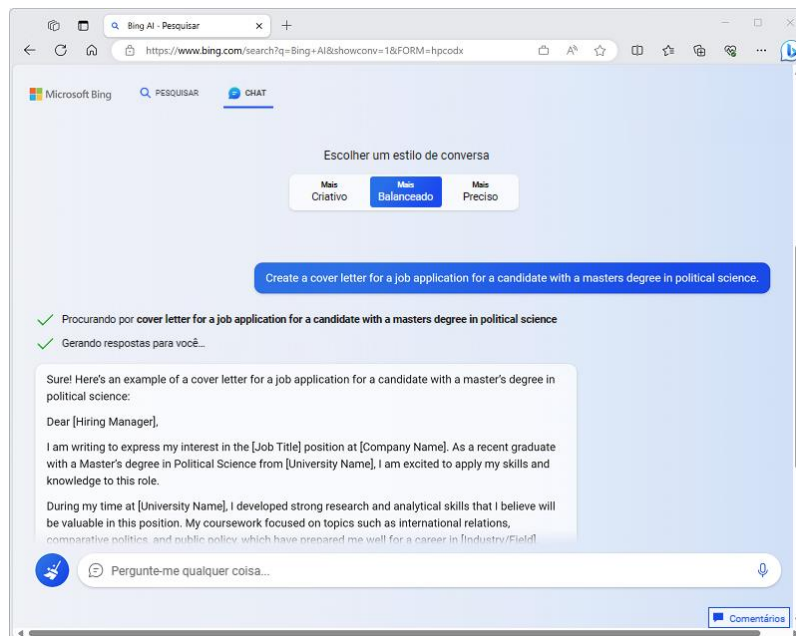
Outros exemplos de copilotos – Navegador Microsoft Edge



Fonte: Microsoft

O que são copilotos?

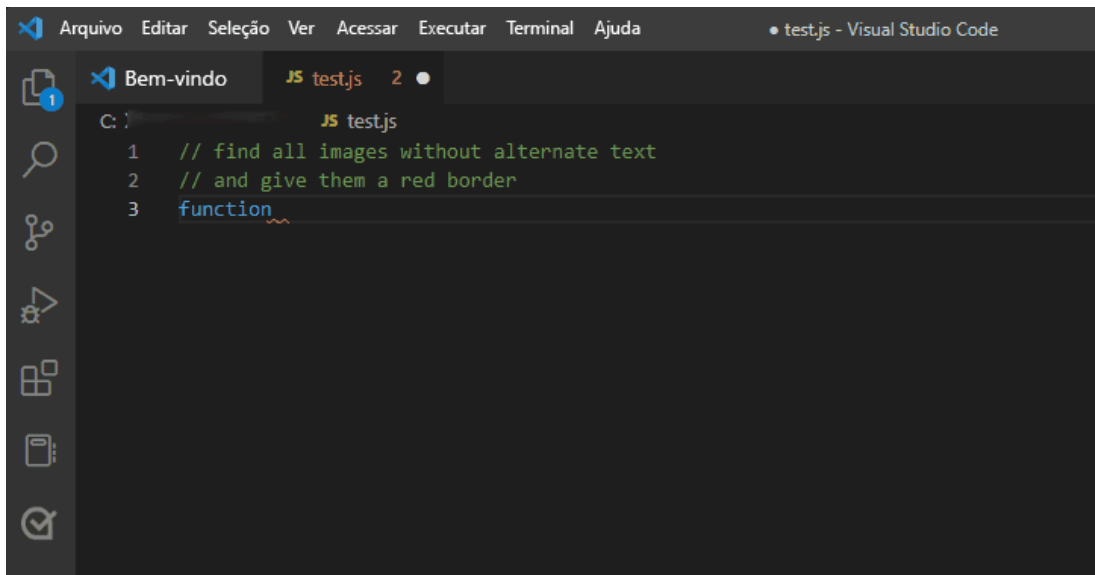
Outros exemplos de copilotos – Microsoft Bing



Fonte: Microsoft

O que são copilotos?

Outros exemplos de copilotos – GitHub Copilot



The screenshot shows the Visual Studio Code interface. The top menu bar includes 'Arquivo', 'Editar', 'Seleção', 'Ver', 'Acessar', 'Executar', 'Terminal', and 'Ajuda'. The title bar indicates the active file is 'test.js - Visual Studio Code'. The left sidebar shows the Explorer view with a file named 'test.js' selected. The main editor area displays the following JavaScript code:

```
C: \> JS test.js
1 // find all images without alternate text
2 // and give them a red border
3 function
```

Fonte: Microsoft

Aprimorar as respostas de IA generativa com a engenharia de prompts

O termo engenharia de prompt descreve o processo de aprimoramento de prompts.

Aprimorar as respostas de IA generativa com a engenharia de prompts

Os desenvolvedores que projetam aplicativos e consumidores que usam aplicativos podem aprimorar a qualidade das respostas da IA gerativa usando linguagem direta, mensagens do sistema, exemplos e/ou dados de fundamentação.

Aprimorar as respostas de IA generativa com a engenharia de prompts

	Descrição	Exemplo
Linguagem direta	Você pode obter conclusões mais úteis sendo explícito sobre o tipo de resposta que deseja.	<i>“Crie uma lista de 10 coisas para fazer em Edimburgo durante o mês de agosto”.</i>

Aprimorar as respostas de IA generativa com a engenharia de prompts

	Descrição	Exemplo
Mensagens do sistema	Descreva como o chat deve funcionar.	"Você é um assistente útil que responde de maneira alegre e amigável ".

Aprimorar as respostas de IA generativa com a engenharia de prompts

	Descrição	Exemplo
Fornecer exemplos	As LLMs geralmente dão suporte ao aprendizado zero-shot no qual as respostas podem ser geradas sem exemplos anteriores.No entanto, você também pode fornecer algumas respostas de exemplo, conhecidas como aprendizado de poucas capturas.	<i>“Visite o castelo pela manhã, antes que as multidões cheguem”.</i>

Aprimorar as respostas de IA generativa com a engenharia de prompts

	Descrição	Exemplo
Dados Básicos	Os prompts podem incluir dados de fundamentação para fornecer contexto.	Incluindo o texto de email com a mensagem “ <i>Resumir meu email</i> ”.

Conceitos básicos do Serviço OpenAI do Azure

O que é o OpenAI do Azure?

- ❑ O **Serviço OpenAI do Azure** é a solução de nuvem da Microsoft para implantar, personalizar e hospedar modelos de linguagem grandes.

O que é o OpenAI do Azure?

O serviço OpenAI do Azure consiste em:

- ❑ Modelos de IA gerativa predefinidos.
- ❑ Funcionalidades de personalização.

O que é o OpenAI do Azure?

- ❑ Ferramentas integradas para detectar e mitigar casos de uso prejudiciais para que os usuários possam implementar a IA com responsabilidade.

O que é o OpenAI do Azure?

- ❑ Segurança corporativa com RBAC (controle de acesso baseado em função) e redes privadas.

O que é o OpenAI do Azure?

Você pode usar vários métodos para desenvolver soluções do Azure OpenAI:

- ☐ Estúdio de IA do Azure,
- ☐ API REST,
- ☐ SDKs com suporte e CLI do Azure.

A quais modelos o OpenAI do Azure dá suporte?

O Azure OpenAI dá suporte a muitos LLMs:	Descrição
GPT-4	Um conjunto de modelos que melhoram o GPT-3.5 e podem compreender e gerar linguagem e código naturais.

A quais modelos o OpenAI do Azure dá suporte?

O Azure OpenAI dá suporte a muitos LLMs:	Descrição
GPT-3.5	Um conjunto de modelos que melhoram o GPT-3 e podem compreender e gerar linguagem e código naturais.

A quais modelos o OpenAI do Azure dá suporte?

O Azure OpenAI dá suporte a muitos LLMs:	Descrição
Incorporações	Um conjunto de modelos que podem converter texto em um formulário de vetor numérico para facilitar a similaridade de texto.

A quais modelos o OpenAI do Azure dá suporte?

O Azure OpenAI dá suporte a muitos LLMs:	Descrição
DALL-E (visualização)	Uma série de modelos em pré-visualização que podem gerar imagens originais a partir de linguagem natural.

Como usar o OpenAI do Azure

Estúdio Azure OpenAI:

- ❑ Crie e implante modelos de IA para aplicativos de software
- ❑ Alimentado por modelos generativos de IA otimizados para diversas tarefas

Como usar o OpenAI do Azure

Estúdio Azure OpenAI:

- ❑ Criar e implantar modelos de IA para aplicativos de software
- ❑ Alimentado por modelos de IA gerativa otimizados para tarefas diversas

Como usar o OpenAI do Azure

Estúdio Azure OpenAI:

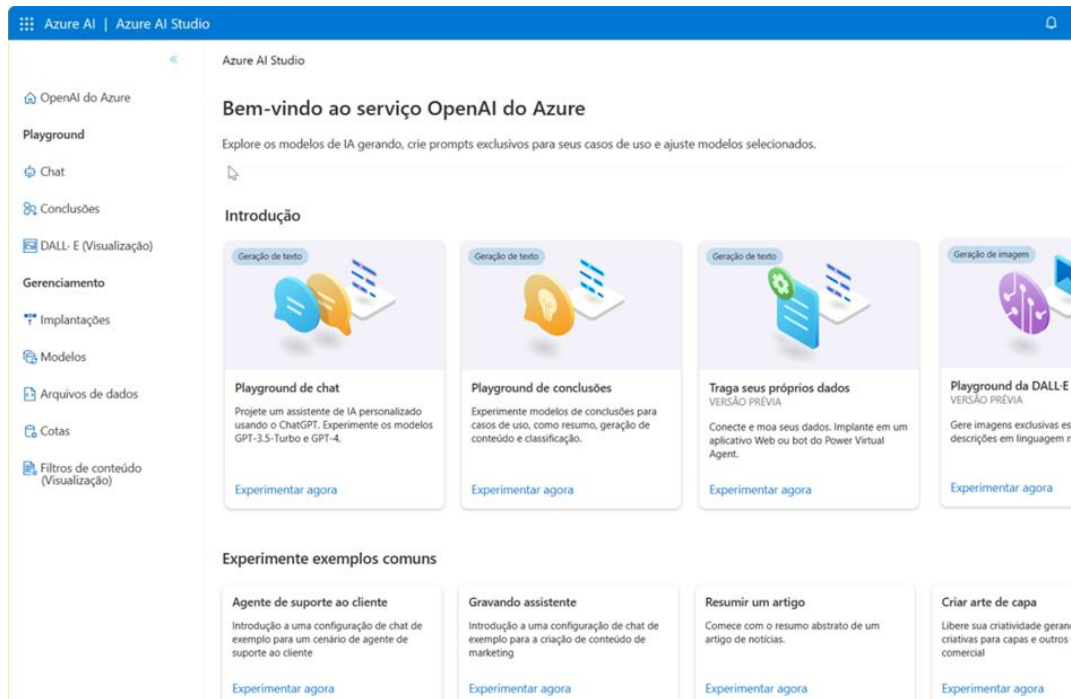
- ❑ Modelos Azure OpenAI incluem: **modelos GPT-4, GPT-3.5, Embeddings e DALL-E**

Como usar o OpenAI do Azure

Playgrounds:

- ❑ Experimente modelos Azure OpenAI sem codificação
- ❑ Use a configuração do assistente para instruir o modelo sobre como ele deve se comportar

Como usar o OpenAI do Azure



The screenshot displays the Azure AI Studio interface. The top navigation bar is blue with the text "Azure AI | Azure AI Studio" and a search icon. The left sidebar contains a menu with the following items: "OpenAI do Azure", "Playground", "Chat", "Conclusões", "DALL-E (Visualização)", "Gerenciamento", "Implantações", "Modelos", "Arquivos de dados", "Cotas", and "Filtros de conteúdo (Visualização)". The main content area is titled "Bem-vindo ao serviço OpenAI do Azure" and includes a subtitle: "Explore os modelos de IA gerando, crie prompts exclusivos para seus casos de uso e ajuste modelos selecionados." Below this is an "Introdução" section with four cards: "Playground de chat" (Projeite um assistente de IA personalizado usando o ChatGPT), "Playground de conclusões" (Experimente modelos de conclusões para casos de uso), "Traga seus próprios dados" (Versão prévia, conecte e mova seus dados), and "Playground da DALL-E" (Versão prévia, gere imagens exclusivas). Each card has an "Experimentar agora" button. At the bottom, there is a section titled "Experimente exemplos comuns" with four cards: "Agente de suporte ao cliente", "Gravando assistente", "Resumir um artigo", and "Criar arte de capa", each with an "Experimentar agora" button.

Fonte: Microsoft

Funcionalidades de linguagem natural do OpenAI do Azure

Os modelos de GPT (transformadores pré-treinados generativos) são excelentes para entender e criar linguagem natural.

Funcionalidades de linguagem natural do OpenAI do Azure

Por exemplo, dado um prompt onde o usuário digita um texto solicitando uma receita culinária.

```
Escreva instruções de receita para uma torta  
de frutas com base nestes ingredientes:
```

```
Strawberries  
Blueberries  
Farinha  
Ovos  
Leite
```

Funcionalidades de linguagem natural do OpenAI do Azure

Exemplo da resposta gerada pelo Azure OpenAI

Instruções:

1. Pré-aqueça seu forno a 350 graus Fahrenheit.
2. Em uma tigela média, misture 1 xícara de farinha e 1 ovo até formar uma massa.
3. Unte um prato de torta de 9 polegadas com manteiga ou spray de cozinha antiaderente.
4. Enrole a massa e coloque-a no prato de torta.
5. Em uma tigela separada, misture 1 xícara de morangos, 1 xícara de mirtilos e 1 xícara de leite.
6. Despeje a mistura de frutas no prato de torta.
7. Gerar

Recursos de geração de código do OpenAI do Azure

Os modelos GPT traduzem linguagem natural ou trechos de código em código.

A geração de código vai além de apenas escrever código a partir de prompts em linguagem natural.

Recursos de geração de código do OpenAI do Azure

Considerando o seguinte código, ele pode gerar testes de unidade, conforme mostrado à seguir:

Python

```
# Python 3
def mult_numbers(a, b):
    return a * b

# Unit test
def
```



Recursos de geração de código do OpenAI do Azure

Python

```
# Python 3
def mult_numbers(a, b):
    return a * b

# Unit test
def test_mult_numbers():
    assert mult_numbers(3, 4) == 12
    assert mult_numbers(0, 10) == 0
    assert mult_numbers(4, 0) == 0

# Unit test
def test_mult_numbers_negative():
    assert mult_numbers(-1, 10) == -10
    assert mult_numbers(10, -1) == -10
```

Funcionalidades de geração de imagem da OpenAI do Azure

- ❑ Os modelos de IA gerativa podem editar e criar imagens. O modelo que funciona com imagens é chamado DALL-E, que dá suporte à criação de imagem, edição de imagem e criação de variações de imagem.

Funcionalidades de geração de imagem da OpenAI do Azure

- ❑ **Geração de imagens:** Com o DALL-E você pode até solicitar uma imagem em um determinado estilo. Os estilos também podem ser usados para edições e variações.

Funcionalidades de geração de imagem da OpenAI do Azure

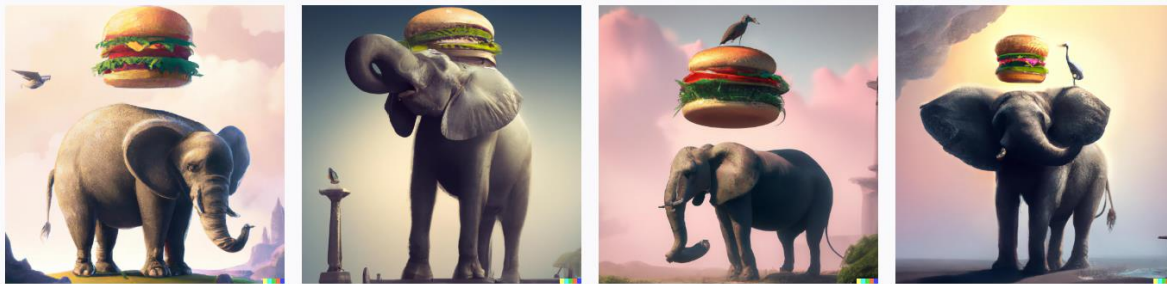
- ❑ **Editando uma imagem:** DALL-E pode editar a imagem conforme solicitado, alterando seu estilo, adicionando ou removendo itens ou gerando novo conteúdo para adicionar.

Funcionalidades de geração de imagem da OpenAI do Azure

- ❑ **Variações de imagem:** variações de imagem podem ser criadas fornecendo uma imagem e especificando quantas variações da imagem você deseja.

Funcionalidades de geração de imagem da OpenAI do Azure

Prompt: “Crie quatro variações da imagem de um elefante com um hambúrguer.”



Fonte: Microsoft

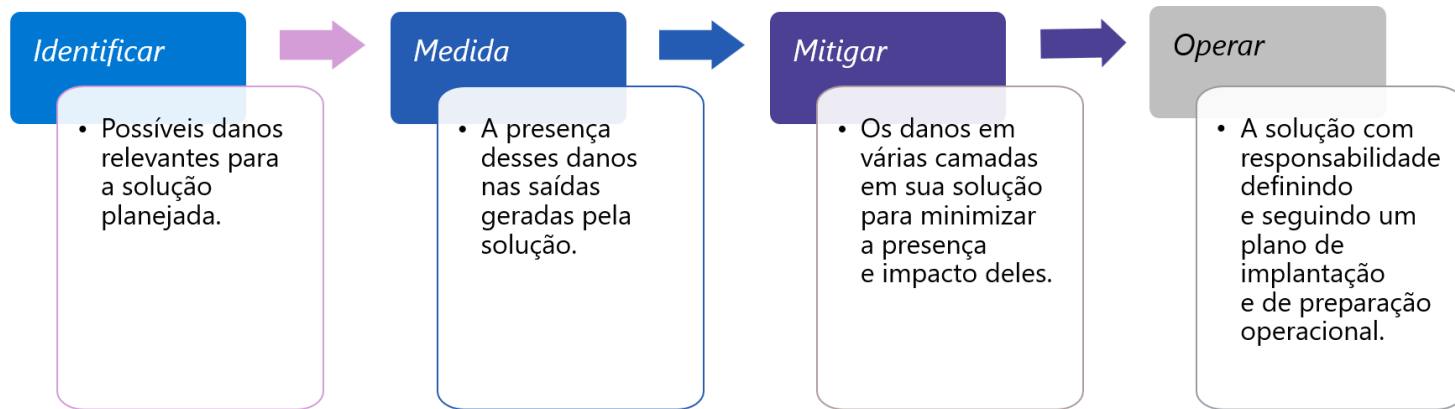
IA generativa responsável

Planejar uma solução de IA generativa responsável

- As quatro fases do processo para desenvolver e implementar um plano de IA responsável são:

Identificar, Medida, Mitigar e Operar

Planejar uma solução de IA generativa responsável



Fonte: Microsoft

Hands On!

***“Falar é fácil.
Mostre-me o código!”***

Linus Torvalds

Links

- <https://aka.ms/ai900-bing-copilot>
- <https://aka.ms/ai900-azure-openai>
- <https://aka.ms/ai900-content-filters>

Links Úteis

- [Documentação Oficial](#)

Dúvidas?

> Fórum/Artigos - <https://web.dio.me/articles>