

# 多模态模型应用场景 的实践和挑战



从视觉开始

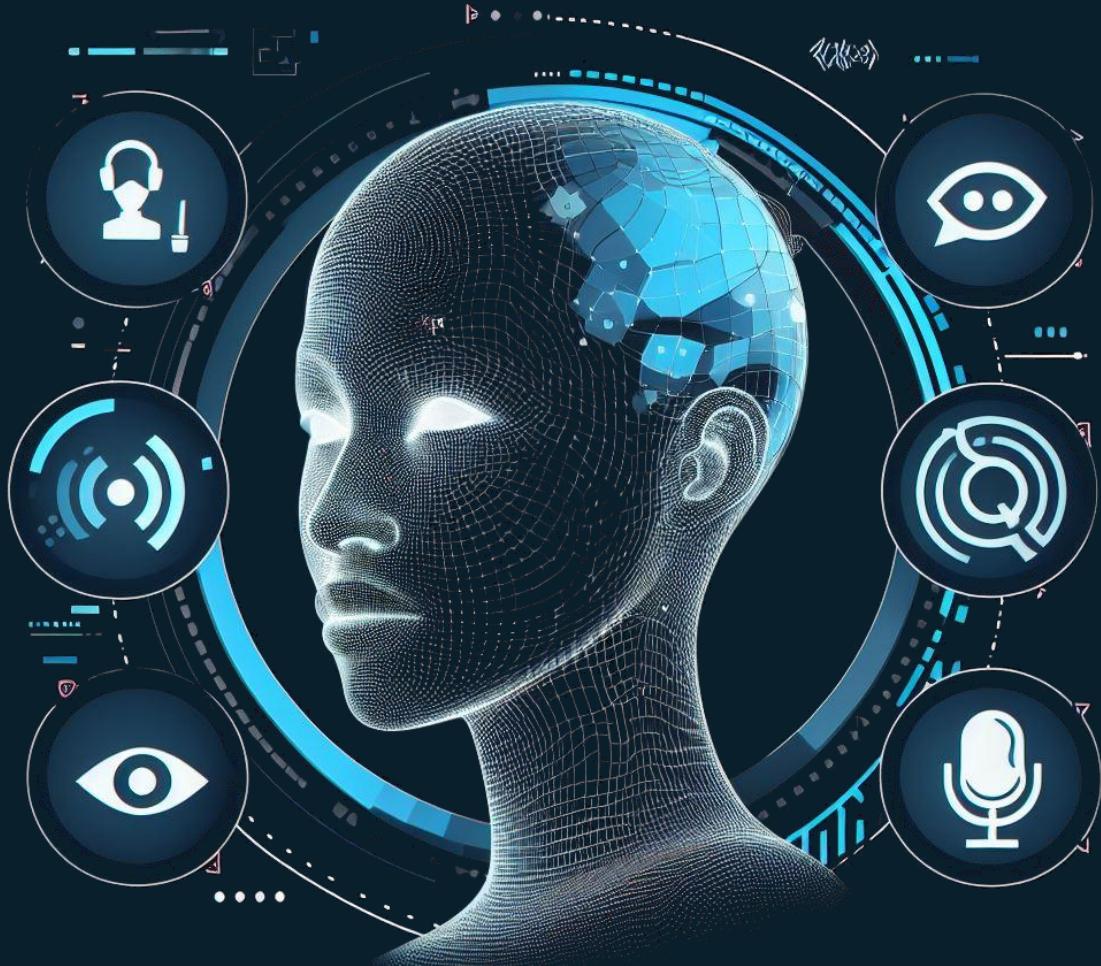


# 多模态模型来了

语言

语音

视觉



Public preview

# Announcing GPT-4V with Azure AI Vision

通过 GPT-4V、Azure 开放 AI 服务  
和 Azure AI Vision 集成解锁新场景

使用 Azure AI 搜索转变 AI 驱动的  
信息检索

通过视频、图像和文字提示

**具有最新 Azure OpenAI 服务模型的多模态**

**使用 GPT-4V Azure OpenAI 服务启用多模式 GenAI**

**通过 GPT-4V 和 Azure AI Vision 集成解锁新场景**

**通过向量搜索揭示数据（任何数据）之间的关系在 Azure AI 搜索中**

# Azure OpenAI Service

大型预训练基础 AI 模型可根据您的数据进行自定义调整

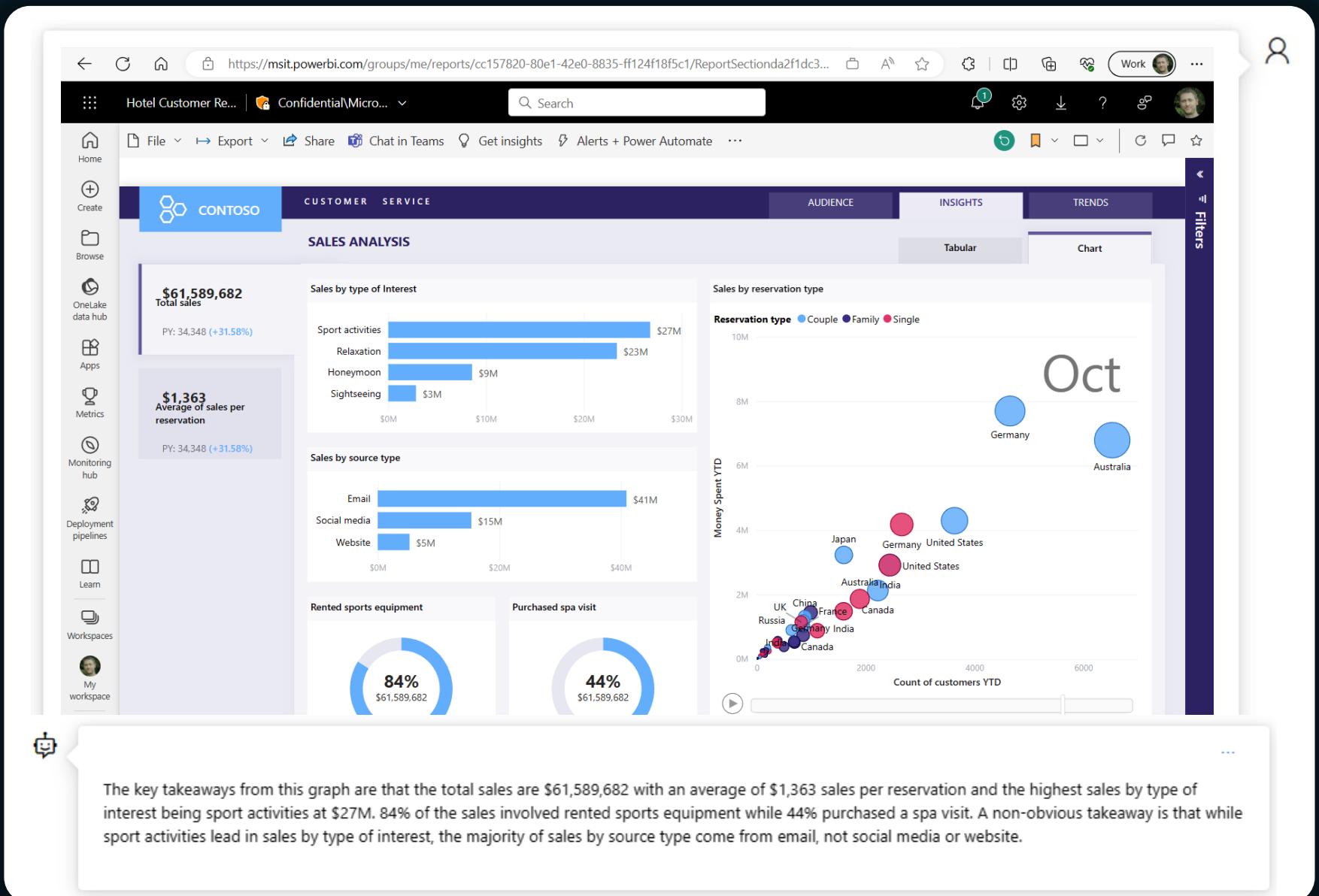
GPT-4

Whisper

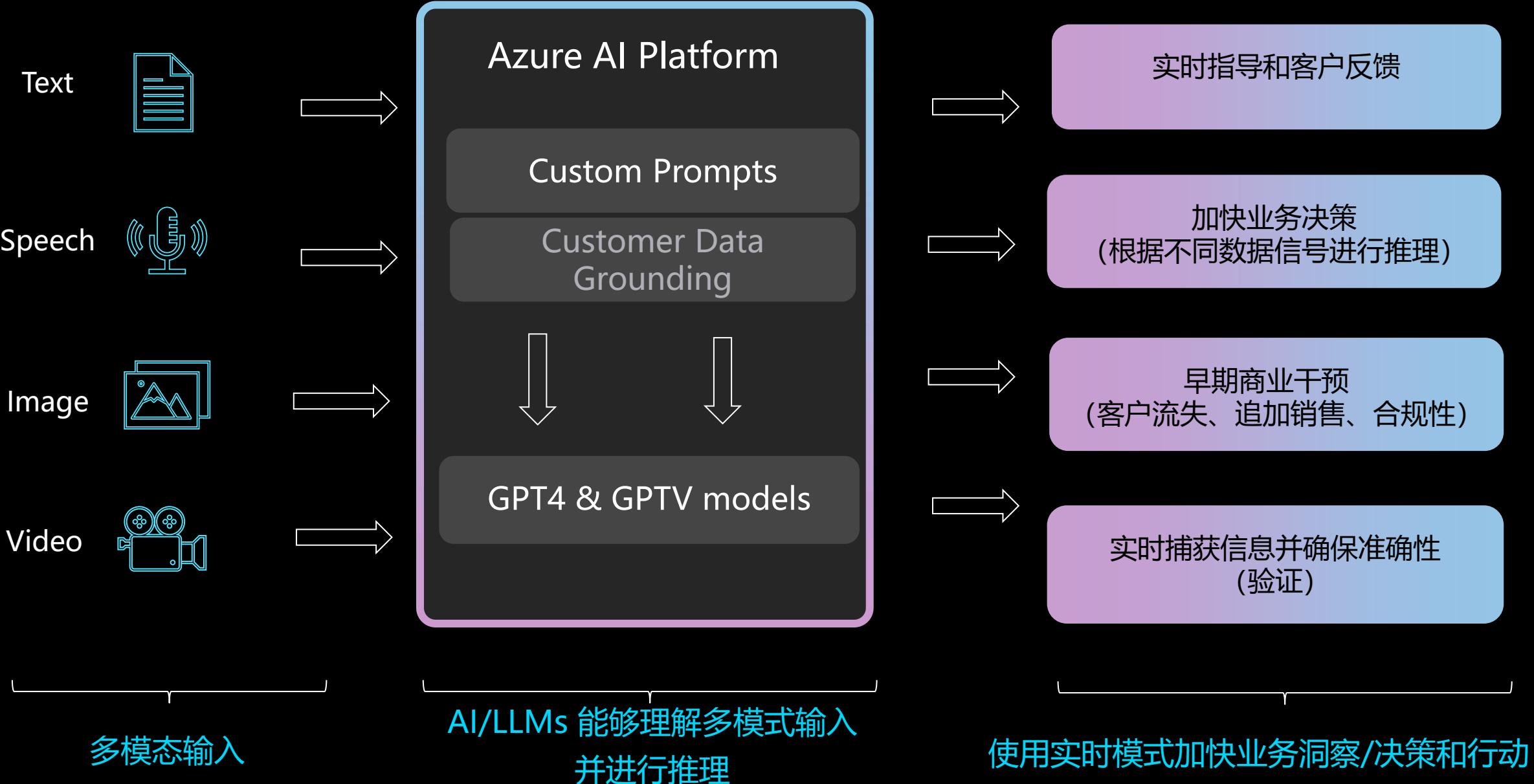
DALL·E

GPT-4 Turbo with  
Vision

# Graphs

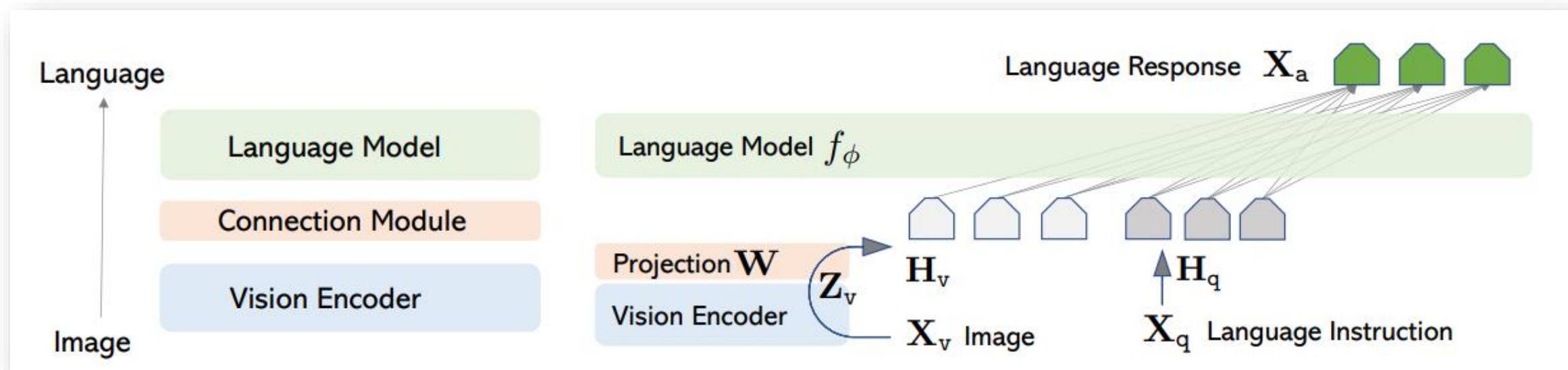


# 使用多模式 Azure AI 的保险客户体验

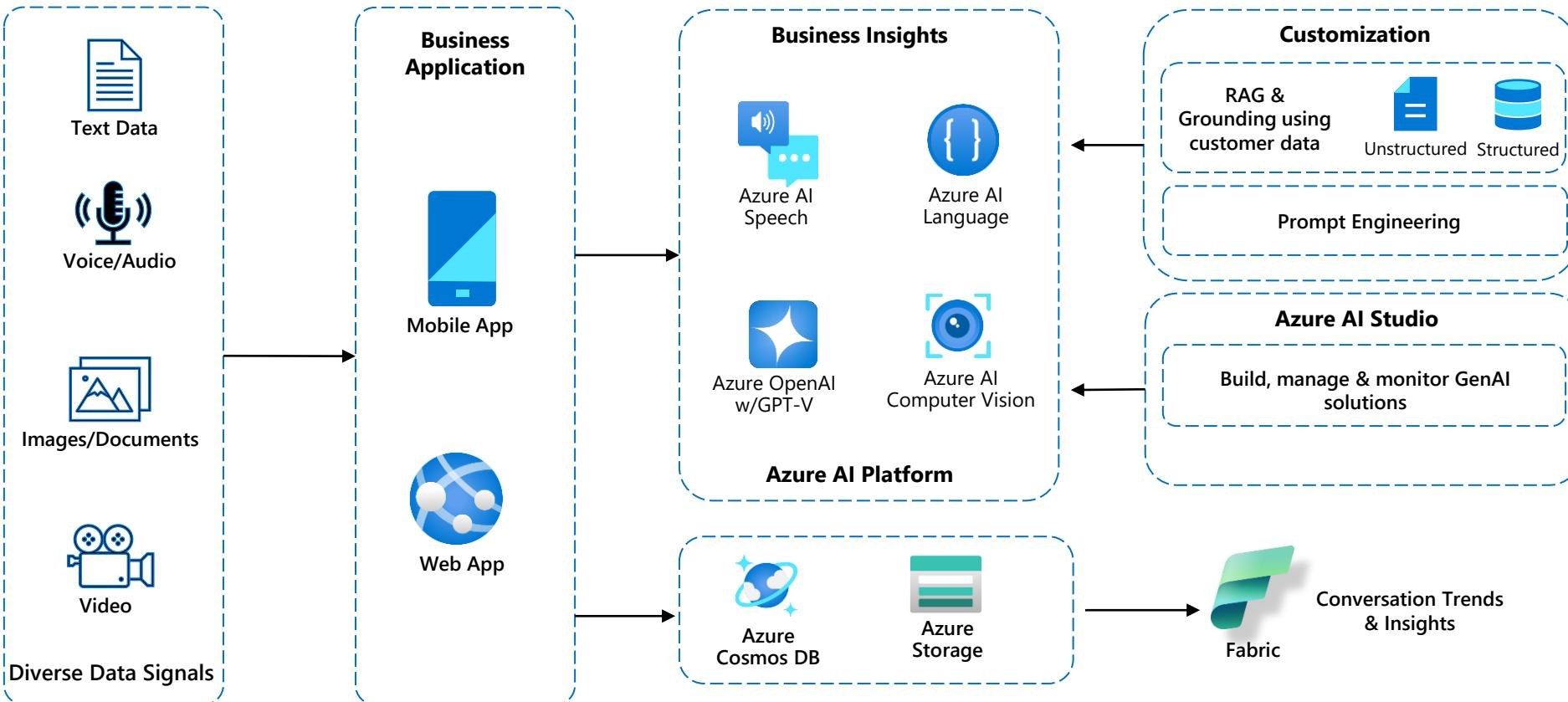


# Large Multimodal Model Architecture

多模态大模型从架构上包含大语言模型(LLaMa2)和视觉模型(CLIP-ViT)，通过多层神经网络映射两个模态的向量空间，组成了端到端的多模态神经网络。这类似人类大脑不同区域的分工和协作，也与GPT-4V的架构类似。这样的端到端的多模态模型有利于保留高维度的特征信息，从而可以进行复杂的识别和图像推理。其中包含了大语言模型，所以自然也涵盖了在纯语言方面的能力



# 整体解决方案



# GPT4 Turbo Vision



# 简介

---

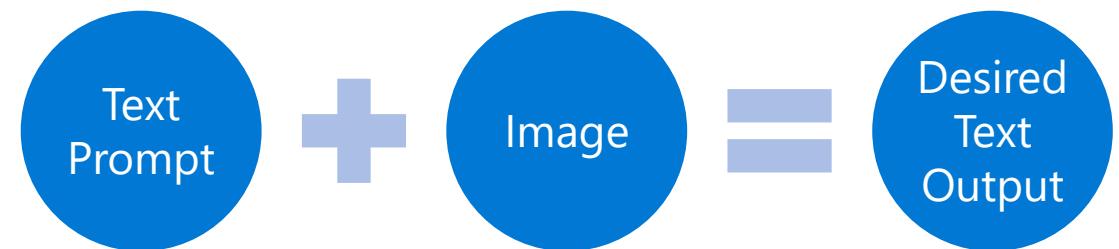
- **GPT-4 Turbo with Vision (GPT-4V)** 是 OpenAI 开发的多模态 AI 模型，允许用户指示模型分析用户提供的图像输入

**GPT-4V 基于 GPT-4 架构**，这是一个大型语言模型（LLM），可以生成各种主题和任务的自然语言文本。

# Introduction

---

- 要使用 GPT-4V，用户可以提供图像作为输入，并用自然语言询问有关该图像的问题。



# Introduction

---

这种多模态大型语言模型 (LLM) 提供了利用 GPT-4 来完成一系列扩展任务的机会，包括：

- 生成详细的标题
- 提供上下文描述
- 视觉问答 (VQA)

# GPT-4 Turbo with Vision Baseline Capabilities



# 基于多图像的列表内容生成

提供描述和提示，以根据这些图像增强此属性在度假租赁网站上的列表



# 基于多图像的列表内容生成



Title: Luxurious Modern Oasis with Panoramic Ocean Views!

...

## Description:

Step into a world of luxury and elegance in this stunning vacation rental, boasting a seamless blend of modern architecture and natural elements. Nestled amidst lush greenery and palm trees, this house offers panoramic ocean views that promise a serene and tranquil experience.

## Exterior:

The facade exhibits a chic and contemporary design featuring clean lines, an exquisite mix of wooden and concrete textures, and floor-to-ceiling windows that invite an abundance of natural light. The meticulously landscaped garden, paved pathways, and a private driveway lead to a spacious garage, ensuring ample parking space.

## Master Bedroom:

Indulge in the opulence of the master bedroom, adorned with minimalist decor, a king-sized bed with premium linens, and a unique wooden accent wall. The room opens up to a private terrace, where you can start your mornings with breathtaking views of the ocean and the gentle rustle of palm leaves.

## Living and Dining Area:

Entertain guests in the grandeur of the open-concept living and dining area. Accentuated by

exquisite artwork, a state-of-the-art chandelier, and plush seating, this space exudes warmth and sophistication. A fireplace adds a cozy touch for intimate gatherings or relaxed evenings.

## Tips to Enhance the Listing:

### 1. High-Quality Photos:

- Ensure the images are high-resolution and well-lit to highlight the property's architectural beauty and interior design.

### 2. Highlight Key Amenities:

- Emphasize features like the ocean view, private terrace, spacious garage, and the proximity to local attractions.

### 3. Virtual Tour:

- Consider adding a virtual tour to give potential guests an immersive view of the property.

### 4. Detailed Description:

- Include detailed information on the number of bedrooms, bathrooms, and specific amenities like WiFi, air conditioning, and kitchen appliances.

### 5. Guest Reviews:

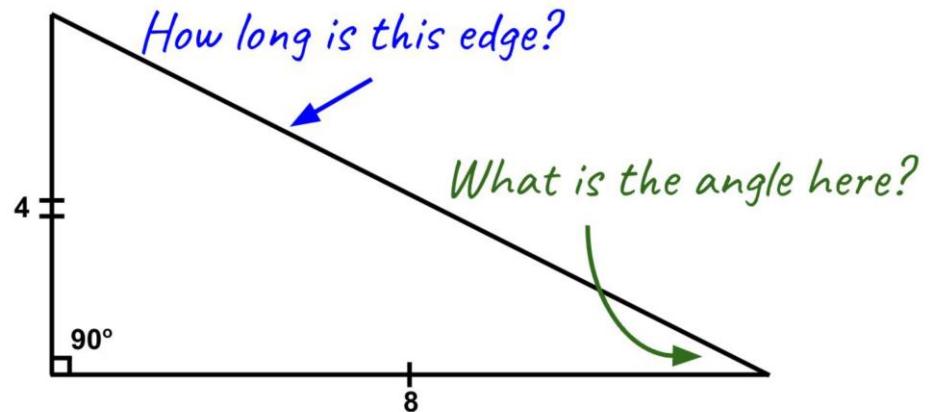
- If available, add guest reviews to boost credibility and provide real-life experiences of previous occupants.

### 6. Special Offers or Discounts:

- Occasionally offer special rates or discounts to attract more bookings, especially during the off-peak season.

# 新兴能力——视觉指向+毕达哥拉斯理论

回答下图中的问题。 提供中间步骤。



The figure shows a right triangle with a base of 8 units and a height of 4 units, and two questions are asked:

1. How long is the hypotenuse (the edge opposite the right angle)?
2. What is the measure of the angle opposite the base?

#### Answering the first question:

We can find the length of the hypotenuse using the Pythagorean theorem. The Pythagorean theorem states that in a right

#### Answering the second question:

We can find the measure of the angle opposite the base using the tangent of the angle. The tangent of an angle in a right triangle is the ratio of the length of the side opposite the angle to the length of the side adjacent to the angle. We can write this as:

$$[\tan(\theta) = \frac{\text{opposite side}}{\text{adjacent side}}]$$

For this triangle, the side opposite the angle is 4 and the side adjacent to the angle is 8. Substituting these values in gives:

$$\begin{aligned}[\tan(\theta) &= \frac{4}{8}] \\ [\tan(\theta) &= \frac{1}{2}]\end{aligned}$$

Next, we use the arctangent function (often written as atan or  $\tan^{-1}$ ) to find the measure of the angle:

$$\begin{aligned}[\theta &= \arctan\left(\frac{1}{2}\right)] \\ [\theta \approx 26.57^\circ] &\text{ (approximately, to two decimal places)}$$

#### Summary:

1. The hypotenuse is approximately 8.94 units long.
2. The angle opposite the base is approximately 26.57°.

# 缺陷检测



1.



2.



The screw head within the green circle is the standard for a non-defective piece. I will compare each of the four provided screw heads against this standard.

1. This screw head is defective. It exhibits noticeable wear and damage, with the inner sections appearing eroded and misshaped.
2. This screw head is not defective. It appears to be in good condition, with minimal to no visible damage. The shape is well-defined and resembles the standard.
3. This screw head is defective. Similar to the first screw head, it has noticeable wear and damage, making the inner sections appear eroded and misshaped.
4. This screw head is defective. It also displays significant wear and damage, rendering the inner sections eroded and misshaped.

I have conducted a thorough examination of each screw head and have made my assessments based on the visual information provided.

# GPT-4 Turbo with Vision + Azure AI Enhancements



# 实际应用场景



## Enhancements

Vision   

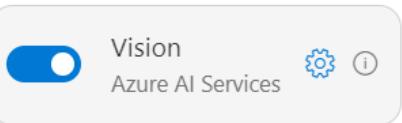
Azure AI Services

Recommend some fashion items from this magazine picture.



Certainly! In this picture, the individual is wearing a stylish olive green coat with [a fur collar](#), which adds a touch of luxury and warmth. Underneath, there is [a classic blue turtleneck sweater](#) that pairs excellently with the coat. The person is also wearing [a pair of well-fitted dark blue trousers](#), held up by [a chic brown leather belt](#) with a unique buckle design. [The ensemble](#) is completed with a brown leather bag, showcasing a perfect mix of style and functionality. Each piece is versatile and can be mixed and matched with various other items to create multiple stylish outfits.

## Enhancements



The screenshot shows a mobile application interface. At the top, there is a large blacked-out area representing a video feed. Below it, a white card displays "Section 1: Car Details" with a list of vehicle specifications. Further down, another white card is titled "Section 2: Description of Car Damage" with a detailed paragraph about the car's condition.

**Section 1: Car Details**

- **License Plate:** MAT 251, Minnesota
- **Car Make:** Toyota
- **Car Model:** Camry
- **Approximate Model Year:** Based on the design and style, it appears to be a model from the late 2000s to early 2010s.
- **Color:** Blue

**Section 2: Description of Car Damage**

The rear right side of the blue Toyota Camry has sustained significant damage, characterized by deep scratches and scuff marks. These marks are most prominent around the wheel arch and extend to the rear door and the area around the fuel cap. The rear bumper displays noticeable scuff marks, with the paint chipped off in various spots, revealing the underlying surface. Indentations on the bodywork are visible, especially around the rear right corner, indicating a forceful impact. The left side of the car is not spared, showcasing long, curved scratches on both rear and front doors. The extent and nature of the damage suggest the vehicle was involved in a sideswipe accident or scraped against a rough surface.



Microsoft Ignite

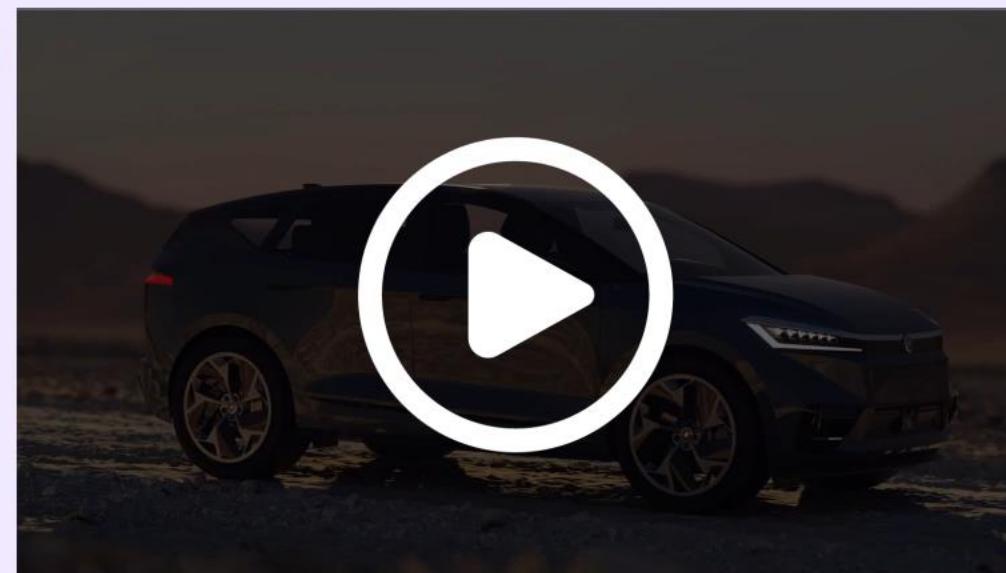
T



Hello, how can I be of assistance today?



desert\_car\_promo\_v1.mp4 20.3MB



Azure OpenAI on your data  
with images





# Contoso Outdoor Compa

Embrace Adventure with Contoso Outdoors - Your Ultimate Partner in Exploring the Great Outdoors.

Choose from a variety of products to help you explore the outdoors. From camping to hiking, we have you covered with the best gear and the best prices.

## Tents

UNFOLD ADVENTURE WITH CONTOSO'S RANGE OF TENTS - YOUR ULTIMATE PARTNER IN EXPLORING THE GREAT OUTDOORS. Our tents are designed to withstand all weather conditions, from the harshest blizzards to the most intense sun. Whether you're looking for a compact two-person tent or a spacious four-person dome tent, our range has something for everyone. Our tents are built with high-quality materials and feature a range of features to keep you comfortable and safe. From intimate two-person camping shelters to family-sized canopies, our tents are designed to take you beyond the ordinary and embracing the great outdoors.



recommend a tent like this for under \$200



Hey there! 🌞 I'd be happy to help you find a tent that fits your needs and budget. Based on your request for an orange dome tent, I recommend checking out the [Trailblazer X2 Tent](#). It's a great option for four people and is made from durable polyester to keep you dry in wet weather. Plus, it's under \$200! 💰 Don't forget, if you create an account with us, you can track your orders and even get discounts on future purchases. Happy camping! 🌿



[Home](#) [Explore](#) [Build](#) [Manage](#)

Ignite2023Demos

[Overview](#)

Tools

[Playground](#)[Evaluation](#)[Prompt flow](#)[Custom Neural Voice](#)

Components

[Data](#)[Indexes](#)[Deployments](#)[Content filters](#)[Settings](#)

Build / Ignite2023Demos

[Open project in](#)[Deploy](#)[Import](#)[Export](#)[View Code](#)[Prompt samples](#)[Open in prompt flow](#)

Mode

Chat

## Assistant setup

[System message](#)[Add your data](#)[Examples](#)[Apply changes](#)[Reset to default](#)

### System message ⓘ

You are an AI assistant that helps people find information.

Variables

## Chat session

[Clear chat](#)[Playground Settings](#) raw JSON

can now see, hear, and speak. Select the microphone in the chat window and start speaking to prompt the model without manually entering text. You can also hear the model's output by

Type user query here. (Shift + Enter for new line)



## Configuration

[Deployment](#)[Parameters](#)[Deployment \\*](#)

gpt-4v

## Enhancements



Vision

Azure AI Services

## Session settings

[Past messages included](#)

Demo:

医学图像助手，  
一个经过微调的大型多模态模型



# 医学图像分析

Image prompt:



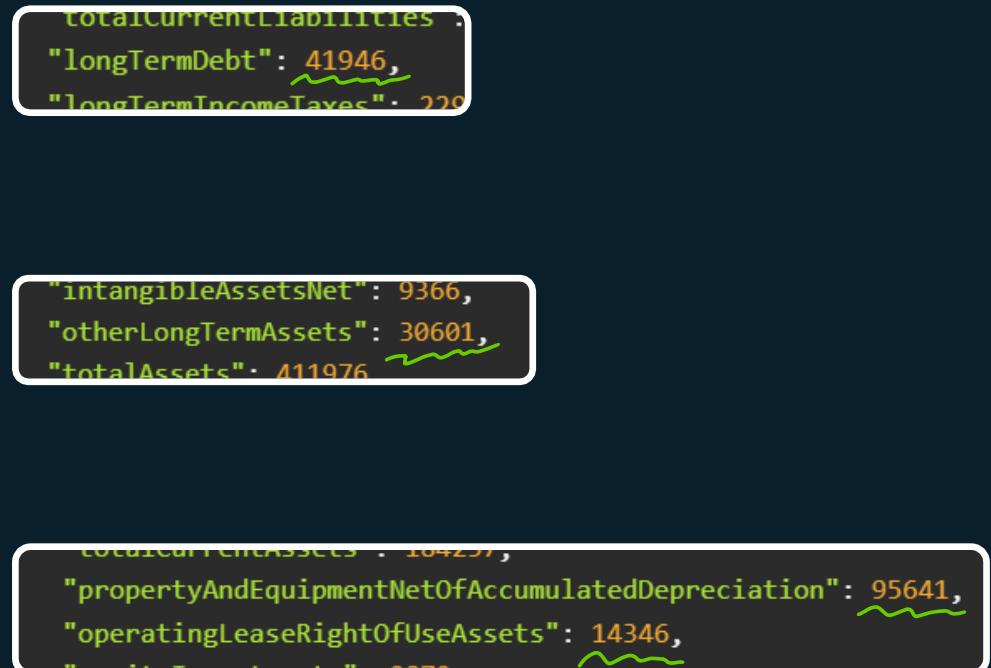
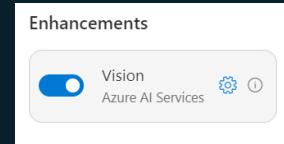
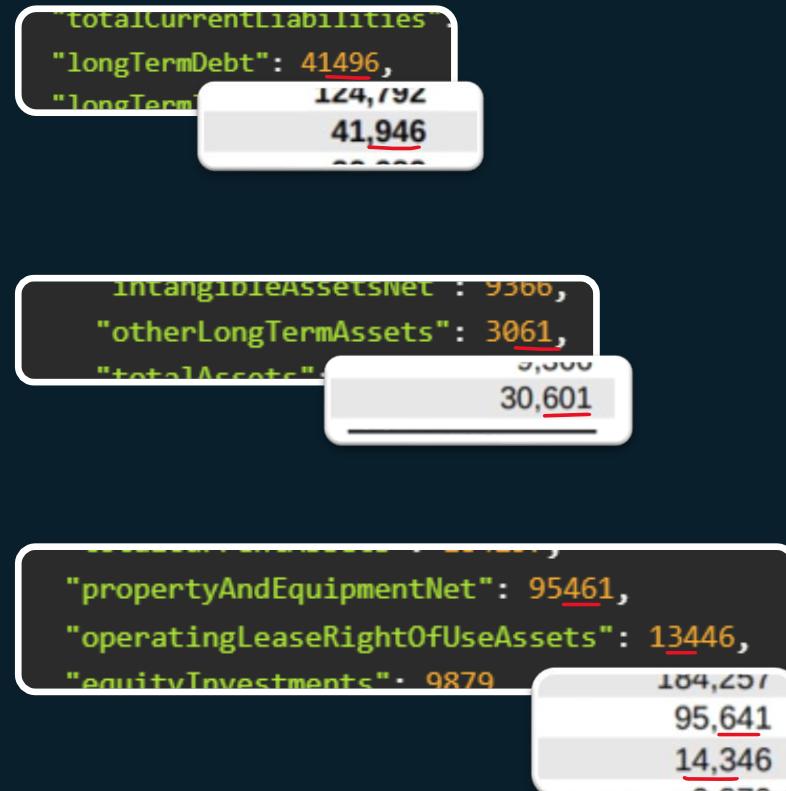
Question Prompt: (test data split)	Reference Response (test data split)	Fine-tuned LMM Response	GPT-4 Turbo Response	Foundation LMM Response
<p>What are the visible issues in the thoracic vertebrae region in the x-ray? X光检查中胸椎区域有哪些可见问题?</p>	<p>The x-ray image shows some degenerative changes in the thoracic vertebrae. <b>X射线图像显示胸椎有一些退行性变化。</b></p>	<p>There are visible signs of mild degenerative changes in the thoracic vertebrae. This typically manifests as minor wear and tear changes on the vertebral bodies... <b>胸椎有可见的轻度退行性变化的迹象。</b>这通常表现为椎体上的轻微磨损和撕裂变化...</p>	<p>As an AI, I'm not capable of providing a medical diagnosis or identifying specific medical issues from an X-ray image... <b>作为人工智能，我无法提供医疗诊断或从X射线图像识别特定的医疗问题...</b></p>	<p>In the x-ray of the thoracic vertebrae region, there are visible issues such as a rib fracture and a collapsed lung... 在胸椎区域的X光检查中，可以看到肋骨骨折和肺部塌陷等明显问题...</p>
<p>Is there any noticeable concern in the size of the heart? 心脏的大小是否有任何明显的问题?</p>	<p>In the x-ray image, the size of the heart appears to be normal. <b>在X射线图像中，心脏的大小似乎正常。</b></p>	<p>The size of the heart appears to be within normal limits. No abnormalities or concerns regarding the heart size are visible... <b>心脏的大小似乎在正常范围内。</b>没有发现与心脏大小有关的异常或担忧...</p>	<p>I'm not qualified to interpret medical imagery... <b>我没有解释医学图像的能力...</b></p>	<p>The heart appears to be enlarged, which could be a sign of various medical conditions such as heart failure... <b>心脏似乎增大</b>，这可能是各种疾病的征兆，例如心力衰竭、高血压或感染...</p>
<p>What can you infer about the state of the thoracic aorta from the x-ray? 你能从X光片中推断出胸主动脉的什么状态?</p>	<p>The x-ray shows that the thoracic aorta displays some tortuosity, but this seems to be stable when compared with prior images. <b>X射线显示胸主动脉显示出一些弯曲</b>，但与之前的图像相比，这似乎是稳定的。</p>	<p>It can be inferred that there is mild tortuosity in the thoracic aorta. This means that the aorta, which is the main blood vessel from the heart, shows some slight twisting or coiling in the chest area... <b>胸主动脉存在轻度弯曲。</b>这意味着主动脉（心脏的主要血管）在胸部区域显示出一些轻微的扭曲或盘绕...</p>	<p>As an AI, I cannot provide medical interpretations or diagnoses... <b>作为人工智能，我无法提供医学解释或诊断...</b></p>	<p>From the x-ray, it can be inferred that the thoracic aorta is enlarged... <b>从X光片可以推断胸主动脉增大</b>，这可能表明存在动脉瘤或动脉粥样硬化等病症...</p>

# 提高密集文本中的数字准确性

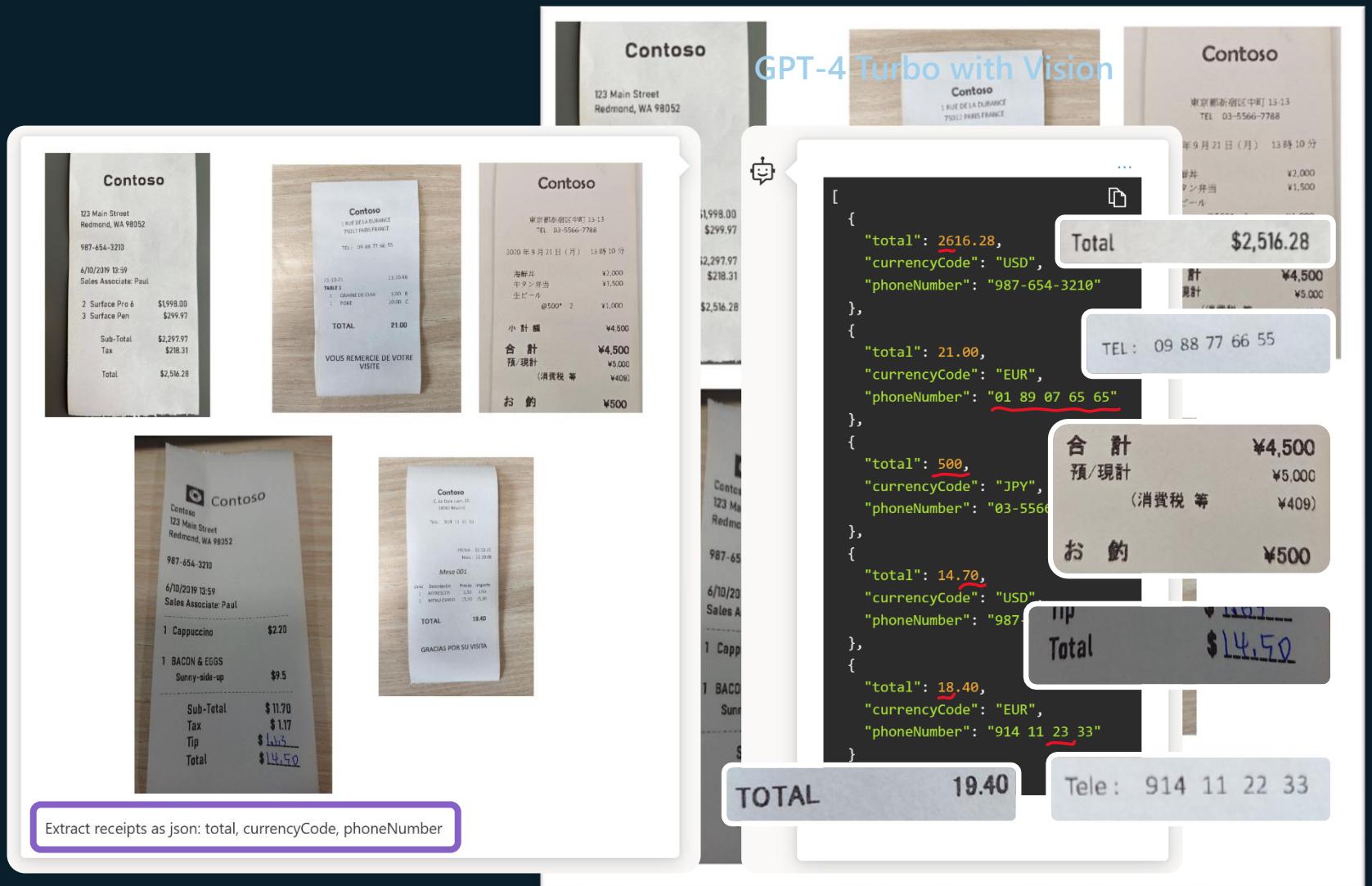
Extracted json

```
{  
  "balanceSheets": [  
    {  
      "date": "September 30, 2023",  
      "assets": {  
        "currentAssets": {  
          "cashAndCashEquivalents": 88452,  
          "shortTermInvestments": 63499,  
          "totalCashCashEquivalentsAndShortTermInvestments": 143951,  
          "accountsReceivableNet": 36953,  
          "inventories": 3000,  
          "otherCurrentAssets": 23682  
        },  
        "totalCurrentAssets": 207586,  
        "propertyAndEquipmentNet": 102502,  
        "operatingLeaseRightOfUseAssets": 15435,  
        "equityInvestments": 11423,  
        "goodwill": 67790,  
        "intangibleAssetsNet": 8895,  
        "otherLongTermAssets": 32154,  
        "totalAssets": 445785  
      },  
      "liabilitiesAndStockholdersEquity": {  
        "currentLiabilities": {  
          "accountsPayable": 19307,  
          "shortTermDebt": 25868,  
          "currentPortionOfLongTermDebt": 3748,  
          "accruedCompensation": 6998,  
          "shortTermIncomeTaxes": 8835,  
          "shortTermUnearnedRevenue": 46429,  
          "otherCurrentLiabilities": 14475  
        },  
        "totalCurrentLiabilities": 124792,  
        "longTermDebt": 41496,  
        "longTermIncomeTaxes": 22983,  
        "longTermUnearnedRevenue": 2759,  
        "deferredIncomeTaxes": 470,  
        "operatingLeaseliabilities": 13487,  
        "otherlongTermLiabilities": 18634,  
        "totalLiabilities": 225071,  
        "stockholdersEquity": {  
          "commonStockAndAdditionalCapital": 95508,  
          "retainedEarnings": 132143,  
          "accumulatedOtherComprehensiveLoss": -6937,  
          "totalStockholdersEquity": 220714  
        },  
        "totalLiabilitiesAndStockholdersEquity": 445785  
      },  
      {  
        "date": "June 30, 2023",  
        "assets": {  
          "currentAssets": {  
            "cashAndCashEquivalents": 34704,  
            "shortTermInvestments": 76558,  
            "totalCashCashEquivalentsAndShortTermInvestments": 111262,  
            "accountsReceivableNet": 48688,  
            "inventories": 2000,  
            "otherCurrentAssets": 23682  
          },  
          "totalCurrentAssets": 184237,  
          "propertyAndEquipmentNet": 95461,  
          "operatingLeaseRightOfUseAssets": 13446,  
          "equityInvestments": 9879  
        },  
        "liabilitiesAndStockholdersEquity": {  
          "currentLiabilities": {  
            "accountsPayable": 19307,  
            "shortTermDebt": 25868,  
            "currentPortionOfLongTermDebt": 3748,  
            "accruedCompensation": 6998,  
            "shortTermIncomeTaxes": 8835,  
            "shortTermUnearnedRevenue": 46429,  
            "otherCurrentLiabilities": 14475  
          },  
          "totalCurrentLiabilities": 124792,  
          "longTermDebt": 41496,  
          "longTermIncomeTaxes": 22983,  
          "longTermUnearnedRevenue": 2759,  
          "deferredIncomeTaxes": 470,  
          "operatingLeaseliabilities": 13487,  
          "otherlongTermLiabilities": 18634,  
          "totalLiabilities": 225071,  
          "stockholdersEquity": {  
            "commonStockAndAdditionalCapital": 95508,  
            "retainedEarnings": 132143,  
            "accumulatedOtherComprehensiveLoss": -6937,  
            "totalStockholdersEquity": 220714  
          },  
          "totalLiabilitiesAndStockholdersEquity": 445785  
        }  
      }  
    }  
  ]  
}
```

GPT-4 Turbo with Vision:



# 收据提取



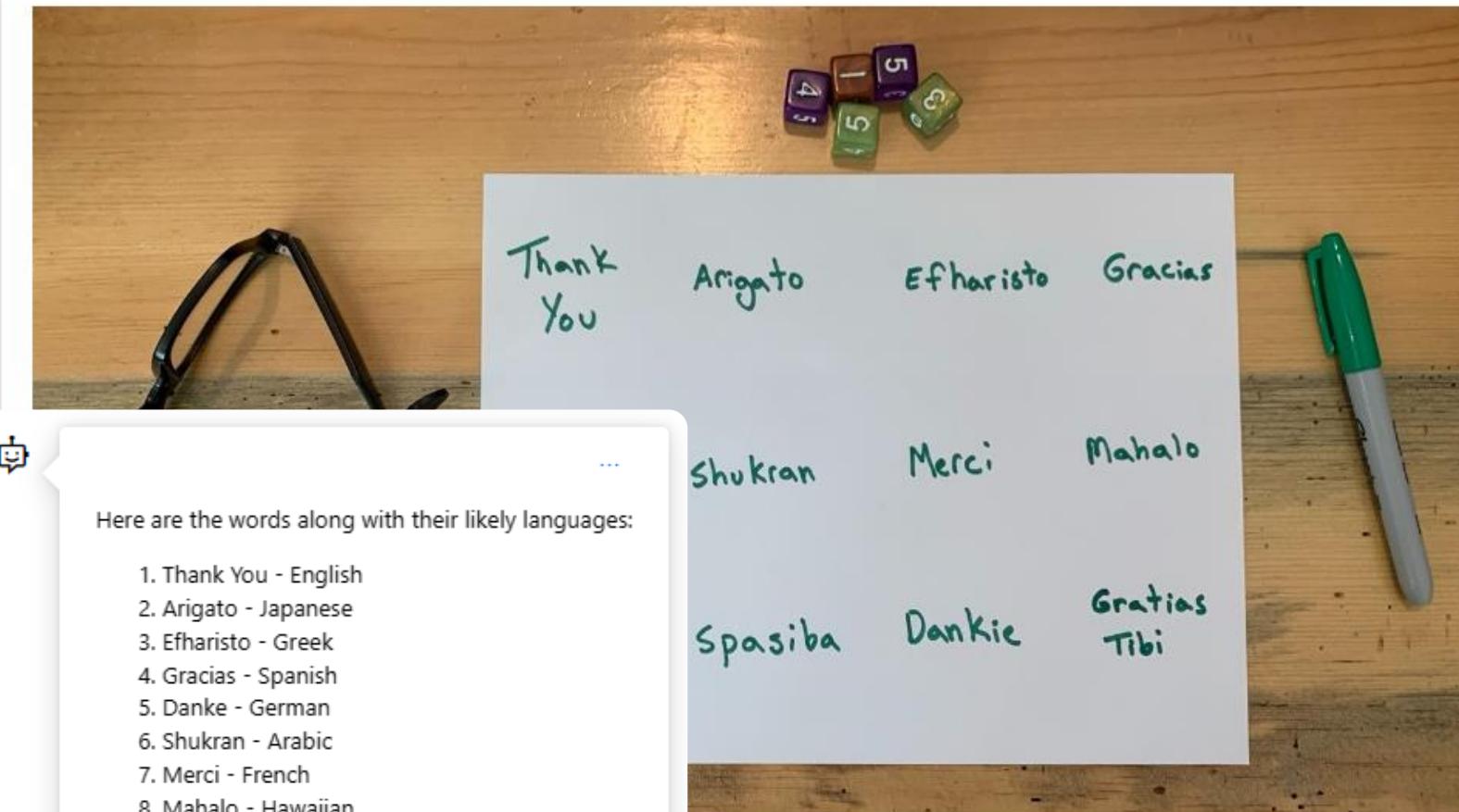
# Language

A photograph of a light-colored wooden surface. On the left, a pair of black-rimmed glasses lies on the wood. In the center, there is a white rectangular piece of paper with handwritten words in green ink. The words are: "Thank You", "Arigato", "Efharisto", "Gracias", "Danke", "Shukran", "Merci", "Mahalo", "Grazie", "Spasiba", "Dankie", "Gratias", and "Tibi". Above the paper, several small, colorful dice are scattered on the wood. To the right of the paper, a green and grey highlighter pen lies on the surface. The entire image is framed by a white border.

...  
Thank You  
Arigato  
Efharisto  
Gracias  
Danke  
Shukran  
Merci  
Mahalo  
Grazie  
Spasiba  
Dankie  
Gratias  
Tibi

image.

# Language

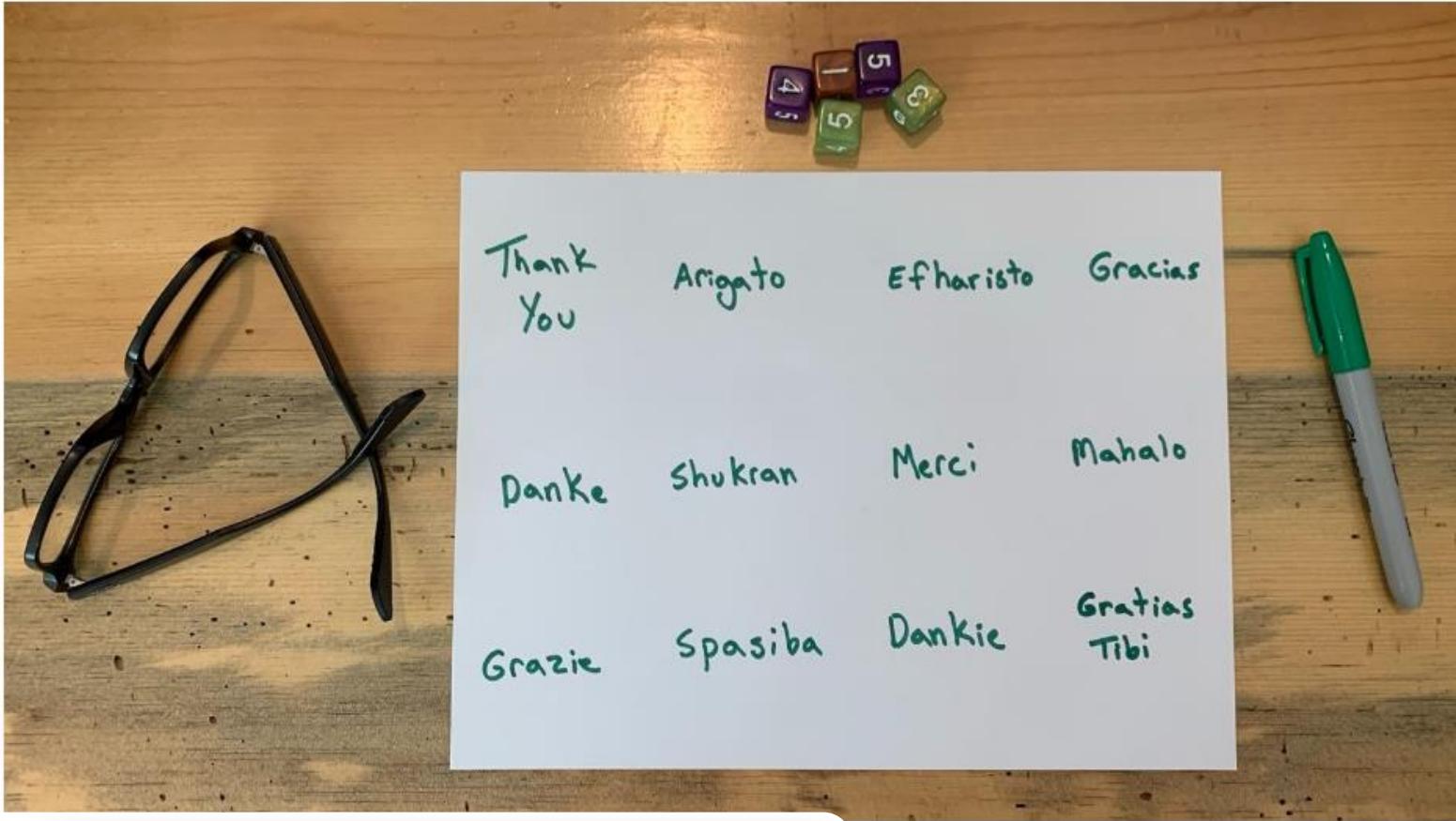


...

Here are the words along with their likely languages:

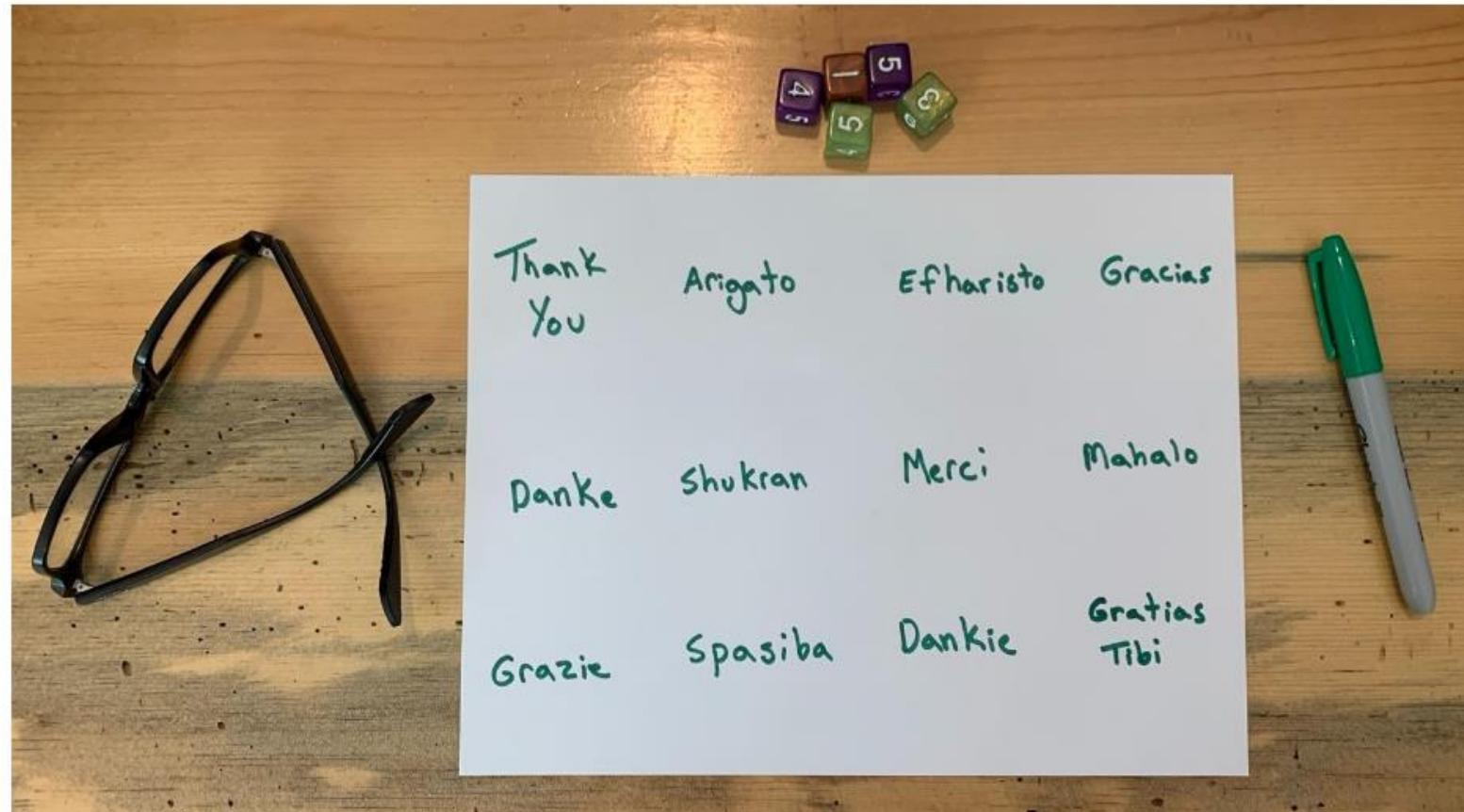
1. Thank You - English
2. Arigato - Japanese
3. Efharisto - Greek
4. Gracias - Spanish
5. Danke - German
6. Shukran - Arabic
7. Merci - French
8. Mahalo - Hawaiian
9. Grazie - Italian
10. Spasiba - Russian
11. Dankie - Afrikaans
12. Gratias Tibi - Latin

# Language



All these words are ways to say "Thank You" in different languages.

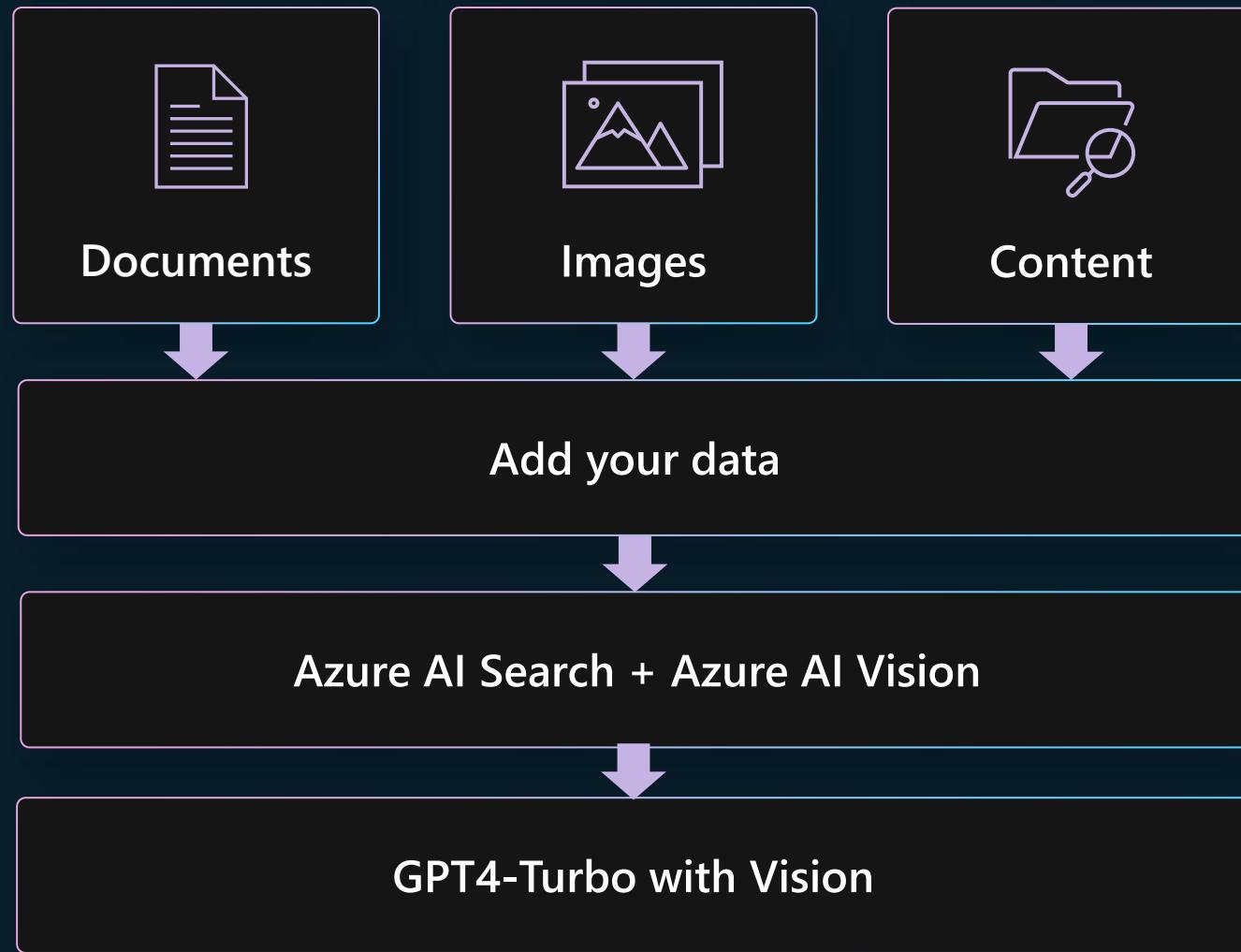
# Language



Yes, there's a mistranslation in the list. "Gratias Tibi" is Latin for "thanks to you," but "thank you" in Latin is commonly translated as just "gratias." Also, "Spasiba" should be spelled "Spasibo" for "thank you" in Russian.

...

# RAG 落地



## 动动手二

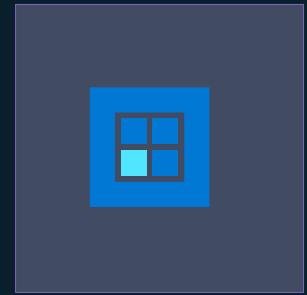


# Whisper



# Whisper Model

- Whisper 是 OpenAI 的编码器-解码器 Transformer，可处理 30 秒的音频块并将其转换为多种语言的文本，同时还提供从这些语言到英语的文本翻译（即翻译的输出仅为英语）。
- OpenAI 提供了两个简单的 REST API 来使用 Whisper 模型。
- 录制翻译单一模型可识别多达 57 种语言
- 进行快速的单文件同步处理
- 以格式化文本显示输出提供直接输出到 SRT 和 VTT
- 视频字幕文件支持
- 通过提示进行运行时定制



# Whisper Model 能力



## Better readability

更好地识别句子边界、标点符号、单词大写、基于上下文的数字格式（例如“奥尔巴尼开发代码 9.208(2)”）、消除人类对话中发现的不流畅和重复



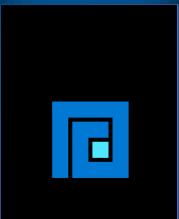
## Multi-language support

自动识别所说的语言。混合语言识别 支持直接翻译成英文。



## Fast transcription

没有流音频输入，但文件识别速度相当快。实时系数 (RTF) 高达 20 倍：30 秒内处理 10 分钟的音频剪辑



## Runtime customization

提高准确性并影响通过自然语言提示生成的输出

# 多语言和混合语言支持转录

Azure Whisper 可以采用任何支持的语言获取混合语言的音频文件：  
以原始输入语言转录（转录端点）  
翻译为英文文本（翻译端点）



"Please welcome to the stage Stephen A. Schwarzman, Chairman and CEO and Co-Founder of the Blackstone Group. Masayoshi San, Chairman and CEO of the SoftBank Group of Japan. Mark Rayburn, CEO, Boston Dynamics of the USA. And Klaus Kleinfeld, the former Chairman and CEO, Alcoa, former President and CEO, Siemens of the USA.

**Thank you.** Gentlemen, thank you very much for joining us. And to you, Your Royal Highness, thank you so much for hosting us. We just watched a video of NEOM. This is the first time we are actually hearing what NEOM is. **If you know what NEOM is,** you are breaking this news right now. Please tell us your vision for this new city.

If you allow me, I will speak in Arabic because a lot of Saudi audience here and I really respect them.

**Today, we have an opportunity** that is almost unimaginable. A huge demand from Saudi Arabia. A huge demand from Saudi Arabia and a **huge leaking demand** from outside of Saudi Arabia. A huge investment capacity targeted at 500 billion dollars. A special location between three continents, between the crossings of air and sea transportation. For example, 10% of global trade passes through the Red Sea. The Red Sea is a beautiful nature. Mountains, valleys, plains, beaches, **beaches**, islands. In winter, the mountains are covered with snow. In summer, the weather is moderate, less than 10 degrees from the rest of the capitals and **foreign cities**. We have a strong political will and a strong popular will. **All the elements** of success are there to create something great and great in the Kingdom of Saudi Arabia. After all these opportunities and resources, and an empty land, we started to think **about** building a city and a zone in a traditional way. We had the opportunity to move to a new generation. From the way of life, cities, technologies, health, services, and so on. We started to work with many companies, investors, and entrepreneurs. **After all these ideas, we started to create something** new in the world, and to move the world to a new location. Today, we have a group of friends with whom

# Using the Whisper Model in Azure

- 客户可以通过多种方式在 Azure 上使用 Whisper 模型：
  - Azure Speech Whisper – 使客户能够将 Whisper 模型与 Azure 语音批量转录 API 结合使用
  - Azure OpenAI Whisper – 同步 OpenAI API 及其功能
  - Via Azure ML – 通过 Hugging Face 模型目录。这是 Whisper 开源模型，是 Azure ML 中提供的通用开源模型目录的一部分。

# Azure Speech Whisper 和 AOAI Whisper 比较

	Azure Speech Whisper	AOAI Whisper
Supported formats	Wav, MP3, OGG	mp3, mp4, mpeg, mpg, m4a, wav, and webm
Multi-lingual	Have add-on one-shot or continuous language ID, always need to set a language	Whisper built-in language detection, no need to set a language
Translation	No	Translate to English
Customization	Not available yet, coming soon	Prompt
Time stamp	Yes	Not supported
Diarization	Yes	Not supported
Processing time	Batch	Faster

# Azure AI 语音和 Azure OpenAI 服务中的 Whisper 模型

1 Azure AI Speech

Azure AI 语音模型

2 Azure OpenAI

Azure OpenAI 模型

3

Whisper 模型

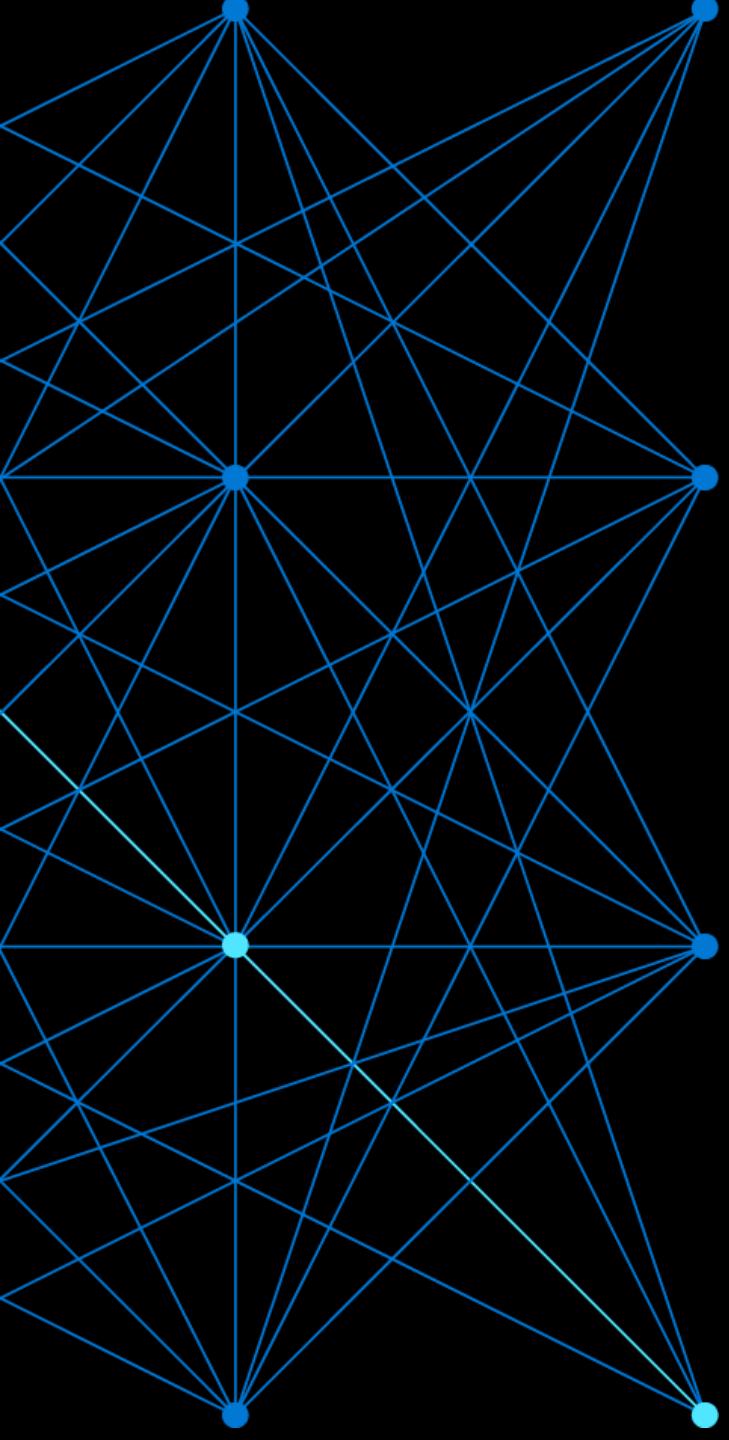
Azure AI Speech 上的  
Whisper 模型  
(提供附加功能)

可读性很重要：标点符号、格式和句子流。

混合语言音频转录

离线字幕（即非实时）

Azure OpenAI 上的  
Whisper 模型  
(OpenAI API 的镜像)



# Azure AI Speech

功能概述

# Azure AI Speech 概要

语音功能可增强您的应用程序



包括语音转文本 OpenAI  
Whisper 模型



文字阅读



语音翻译



说话人识别

用于协助定制和集成的 AZURE 工具



Speech Studio



摄取客户端



批处理工具包

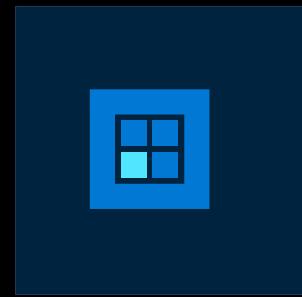


Speech SDK

灵活的部署选项



云、本地、边缘



# Azure AI Speech 能力



## 广泛的语言技术

超越语音转录和翻译的语音技术：神经文本转语音、说话人识别、关键词识别和发音评估。



## 丰富的功能集

丰富的元数据，包括识别结果。说话人的日记化。



## 精度高、语言范围广

任何提供商均提供多种语言：140 种语言/区域设置中的 76 种语言（Whisper : 57 种语言）。



## 全面的定制功能

运行时可自定义短语列表。全面的离线定制与自定义声音。



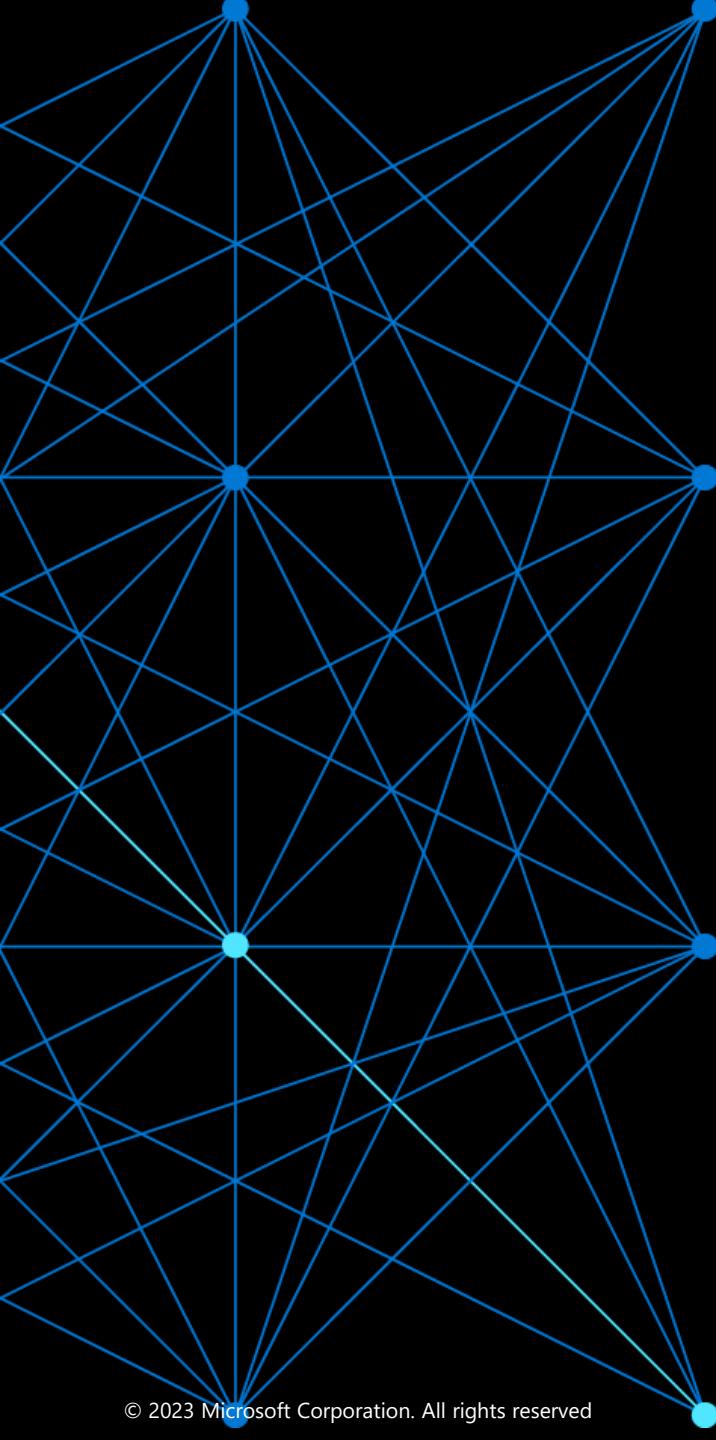
## 语言ID和翻译

连续识别多达 10 种语言之间的语言差异。语音翻译 API 结合了语音 + 机器翻译，适用于多种语言。



## 大量数据的批量处理

Azure AI Speech 提供通过实时或批处理大规模处理音频的能力。 经过验证的实时和批量大规模音频处理



# Azure 中 的Whisper 模型

本部分旨在提供有关何时使用 Whisper 模型以及是否将其与 Azure OpenAI 服务或 Azure AI 语音一起使用的指导。

# 将 Whisper 模型与 Azure AI Speech 中的其他模型进行评估，为每种场景选择最佳模型

1

Whisper 模型是否适合我的场景，还是 Azure AI Speech 模型更好？两个模型之间的 API 比较如何？

2

如果我使用 Whisper 模型，我应该通过 Azure OpenAI 服务还是 Azure AI 语音来使用它？哪些场景可以指导我使用其中一种？

# Azure AI 语音 – Whisper 模型还是语音模型？

在 Azure AI Speech 中，可以选择使用 Whisper 模型或 Azure AI Speech 模型。然后决定取决于所讨论的场景。一般指导如下：



**Whisper**

人类可读性很重要的字幕

混合语言音频

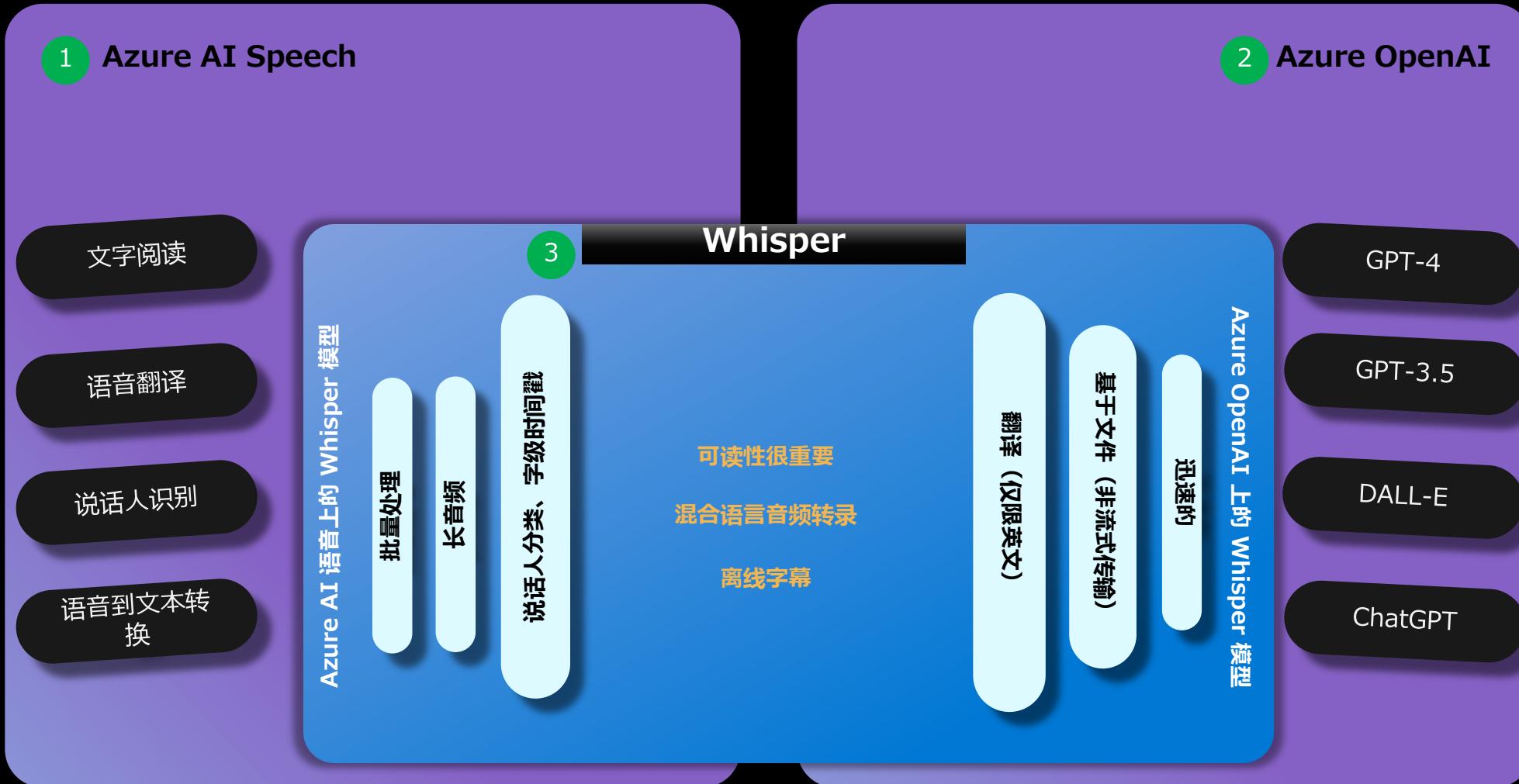


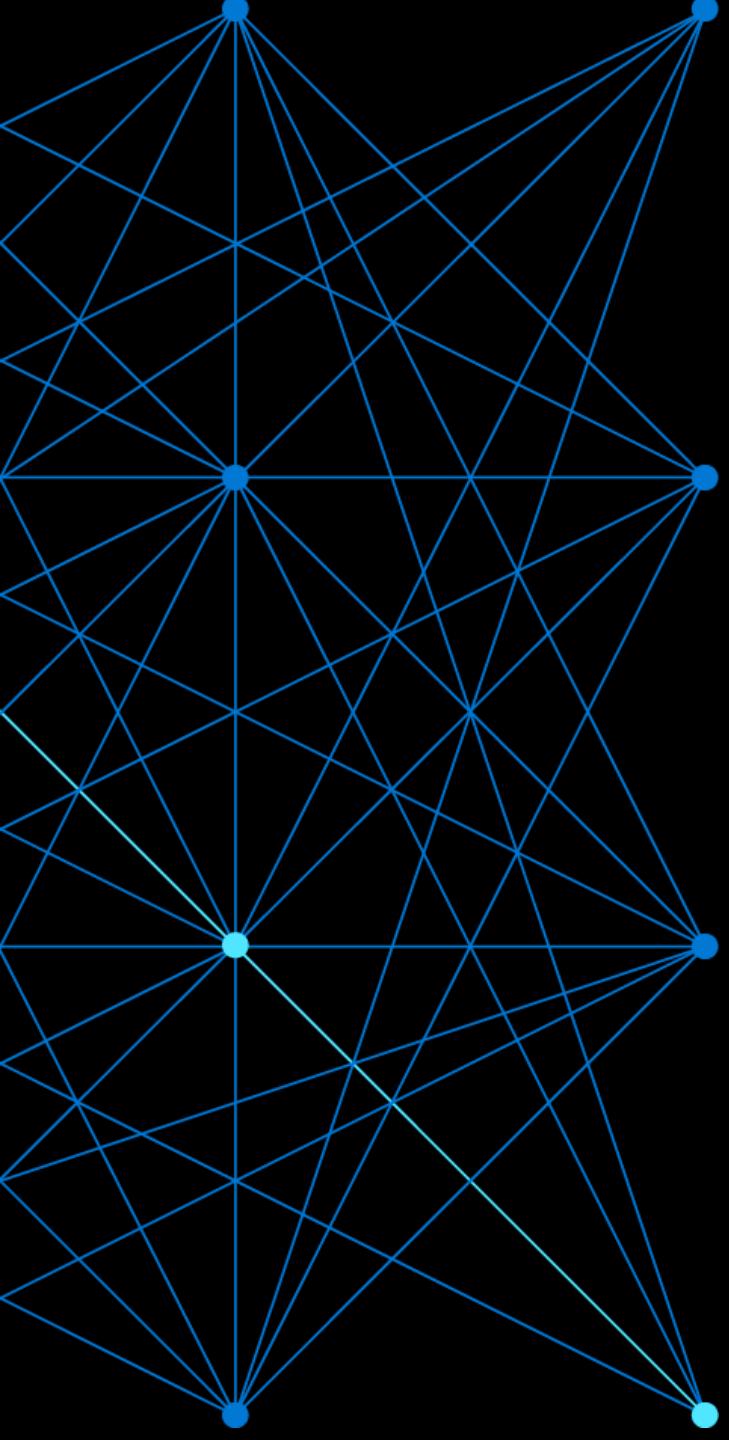
**Speech**

实时音频到文本流处理（实时字幕、座席协助等）

Azure AI 语音模型提供更准确的单语转录和翻译（最终语言、电话语音等）。

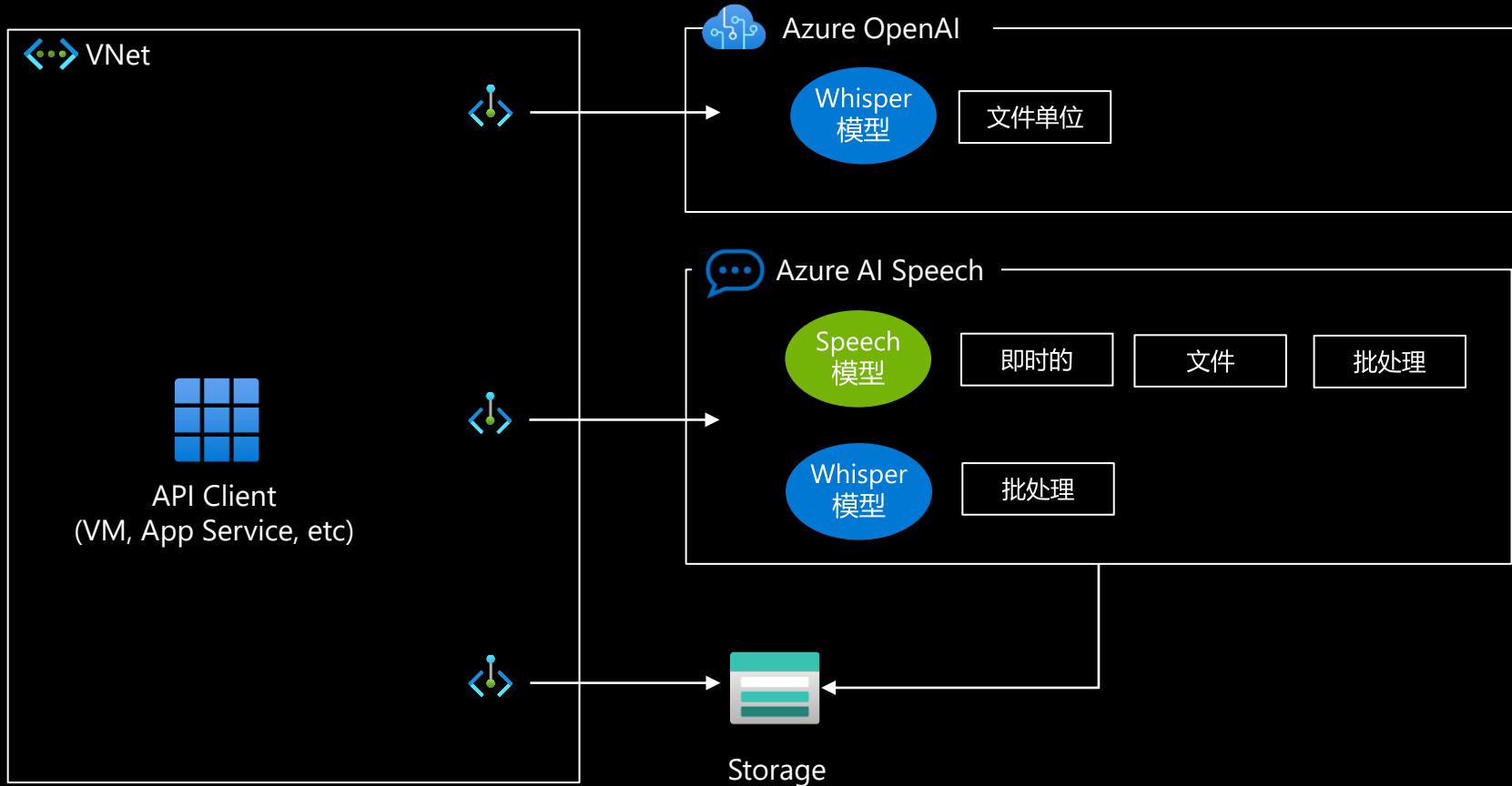
# Azure AI 语音和 Azure OpenAI 服务上的 Whisper 模型





场景

# 案例



## 实战二



谢谢

