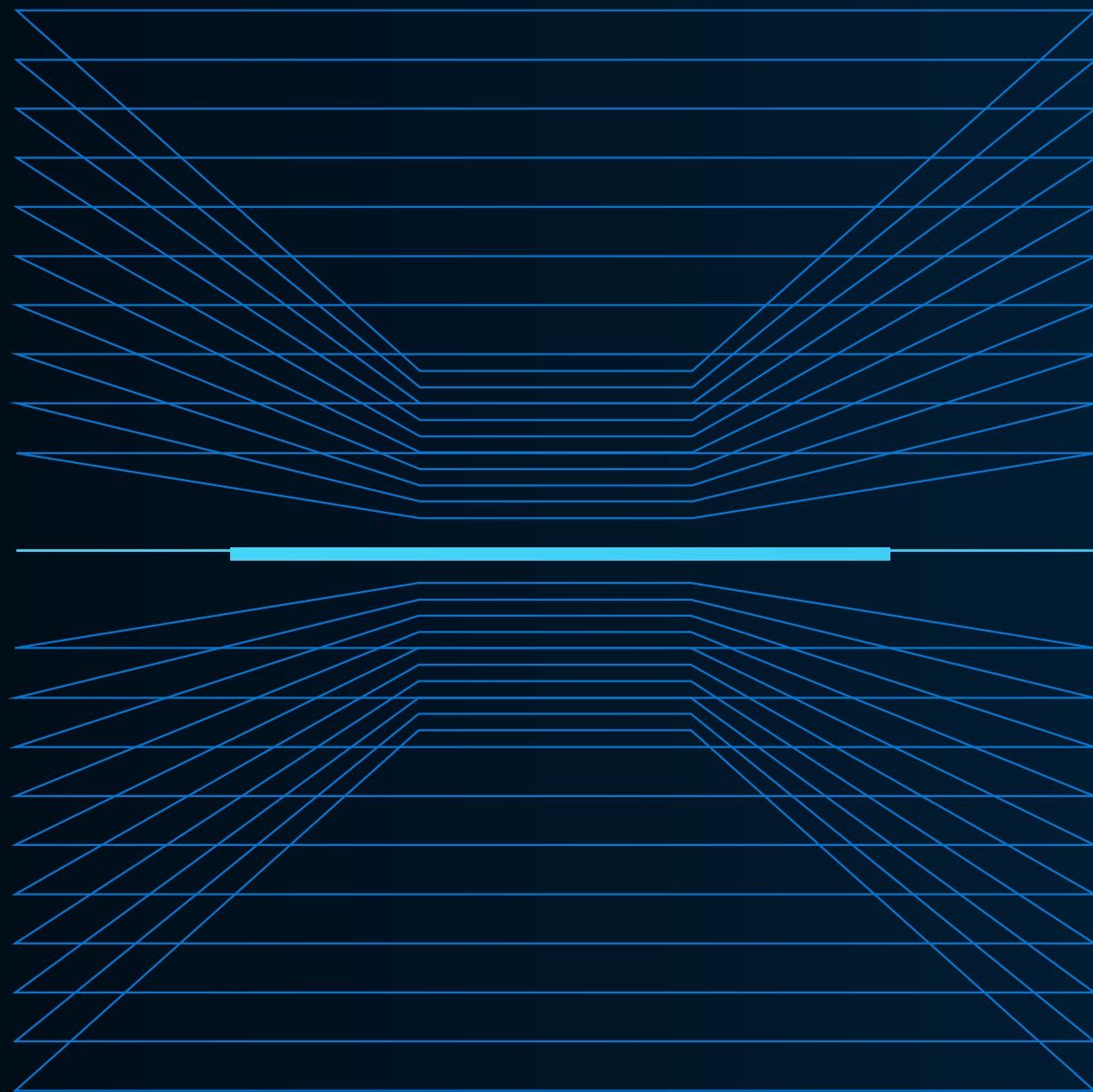




面向企业的 ChatGPT服务介绍

Rade Zheng – 郑柱镭
微软资深云解决方案架构师



Agenda



什么是 ChatGPT



GPT大模型的核心概念



OpenAI 模型能力



Azure OpenAI 企业级能力



微软内的用例场景

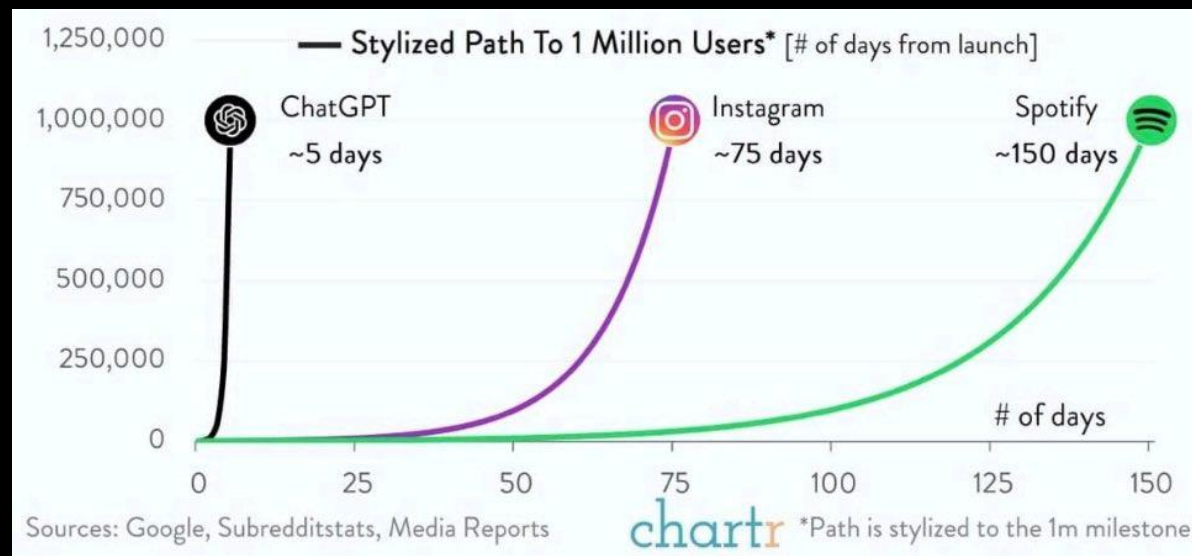
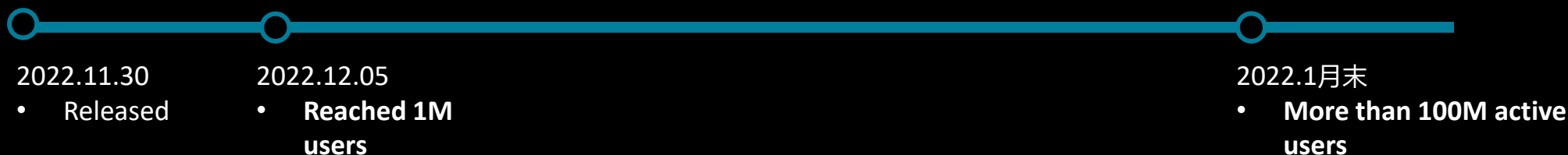


01

什么是ChatGPT

ChatGPT火爆出圈

- ChatGPT 简单理解为专为聊天而生的大规模语言模型
- 5天内100万用户注册，2个月内用户量突破1亿



 Linas Belius [in](#)
@linas.belius

Time it took to reach **1 million users**:

- Netflix - 3.5 years
- Airbnb - 2.5 years
- Facebook - 10 months
- Spotify - 5 months
- Instagram - 2.5 months
- iPhone - 74 days
- ChatGPT - 5 days

ChatGPT is one of those rare moments in technology that **will reshape everything** going forward.

ChatGPT/GPT-4 火爆出圈 - OpenAI 推出的聊天机器人

×

...

GPT-4通过律师资格考试：排名高于90%人类考生

原创 AIGC开放社区 AIGC开放社区

2023-03-21 07:41 发表于河北

收录于合集

#ChatGPT 197 #AIGC 209 #生成式AI 158

#大语言模型 96

专注AIGC领域的专业社区，关注GPT-3、百度文心一言等大语言模型（LLM）的发展和落地，以及国内LLM的发展和落地，欢迎关注！

AIGCOPEN

AIGC开放社区

专注AIGC（生成式人工智能）领域的...

98篇原创内容

公众号

美国伊利诺伊理工大学-芝加哥肯特法学院在官网发布了消息，GPT-4通过了统一律师资格考试。该考试一共7项，GPT-4在民事诉讼、合同法、刑法、物权法、证据法5个学科考试中，得分率高于人类考生。

×

AI能力站 >

...

十大可能被取代的职业

媒体网站insider在与专家交谈后和研究后，得出10个工作被人工智能取代的风险最高

技术岗：程序员、工程师、数据分析师

媒体岗：广告、内容创作、写作、新闻

法律岗：法律或律师助理

市场研究分析师

教师

金融岗：金融分析师、个人财务顾问

交易员

平面设计师

会计师

客服人员

×

...

重磅！网梯睿学AI学习助手：基于ChatGPT的智能化学习新体验

原创 小梯 网梯科技 2023-03-21 18:48

发表于北京

当下，AI技术越来越成熟，其在各个领域的应用也日益广泛。而在这其中，自然语言处理技术无疑是应用最为广泛的，ChatGPT作为其中的佼佼者，更是备受关注和追捧。作为一种语言生成模型，ChatGPT凭借其强大的生成能力和多样的应用场景，迅速成为了广受欢迎的AI应用之一。

2023年3月，网梯睿学通过Microsoft Azure OpenAI直接引入ChatGPT的基础能力，重磅推出【睿学AI学习助手】，为广大学子提供全新的学习方式和学术支持！通过将网梯公司的专业技术和ChatGPT的API相结合，学习助手不仅能够解决用户的问题，还能够根据用户的需求提供个性化的服务和支持，帮助用户更快地掌握知

×

...

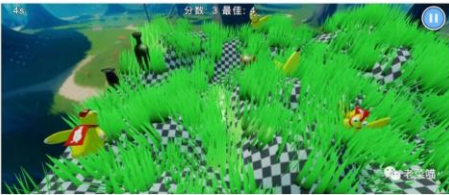
ChatGPT做了一个二次元游戏！

原创 老菜喵 老菜喵 2023-04-07 20:16

发表于广东

1.0 游戏策划设计

继上次借助ChatGPT做了一个3D小游戏后，很多朋友问我，AI可以做大型项目么？还是仅限于简单的小游戏。



*AI生成的3D小游戏

所以二喵准备接着用AI设计一款中型体量的卡牌游戏，发布到微信小游戏和海外平台。

从以下几点着手，来探索AI能力的上限。

1. 策划能力 AI是否能提供完整且有趣的游戏策划案？

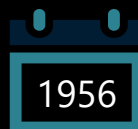
2. 美术能力 AI是否能提供游戏的美术素材？

人工智能

机器学习

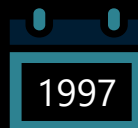
深度学习

生成式 AI



人工智能

计算机科学领域，旨在创造能够复制或超越人类智能的智能机器



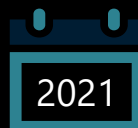
机器学习

人工智能的子集，使机器能够从现有数据中学习，并改进该数据以做出决策或预测



深度学习

一种使用多层神经网络处理数据并做出决策的机器学习技术

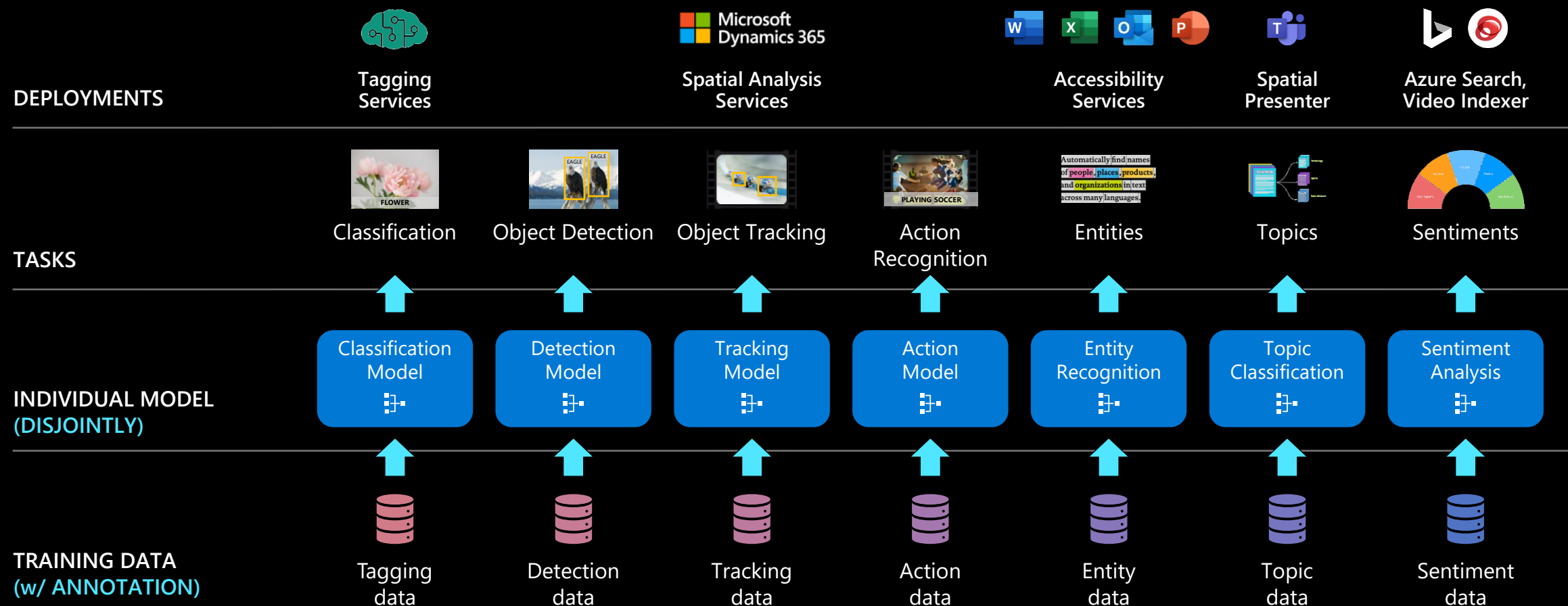


生成式 AI

根据提示或现有数据创建新的书面、视觉和听觉内容

传统模型开发

成本高，部署速度慢 — 每个服务都进行不连贯的训练



GPT模型

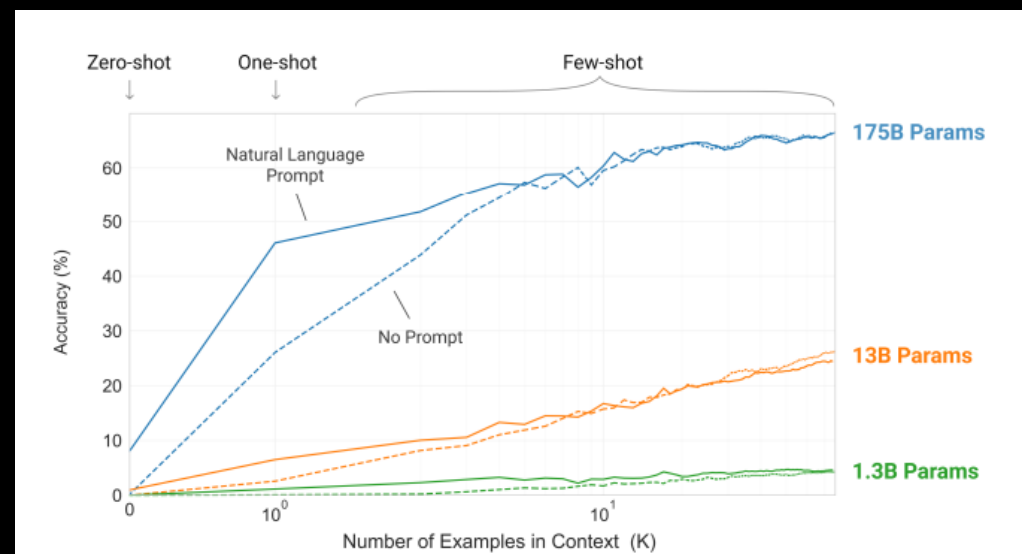
Generative Pre-trained Transformer (GPT) 系列是由OpenAI提出的非常强大的预训练语言模型，这一系列的模型可以在非常复杂的NLP任务中取得非常惊艳的效果，例如文章生成，代码生成，机器翻译，Q&A等，而完成这些任务并不需要有监督学习进行模型微调。而对于一个新的任务，GPT仅仅需要非常少的数据便可以理解这个任务的需求并达到接近或者超过state-of-the-art的方法。

模型	发布时间	参数量	预训练数据量
GPT	2018 年 6 月	1.17 亿	约 5GB
GPT-2	2019 年 2 月	15 亿	40GB
GPT-3	2020 年 5 月	1,750 亿	45TB

GPT-3/4

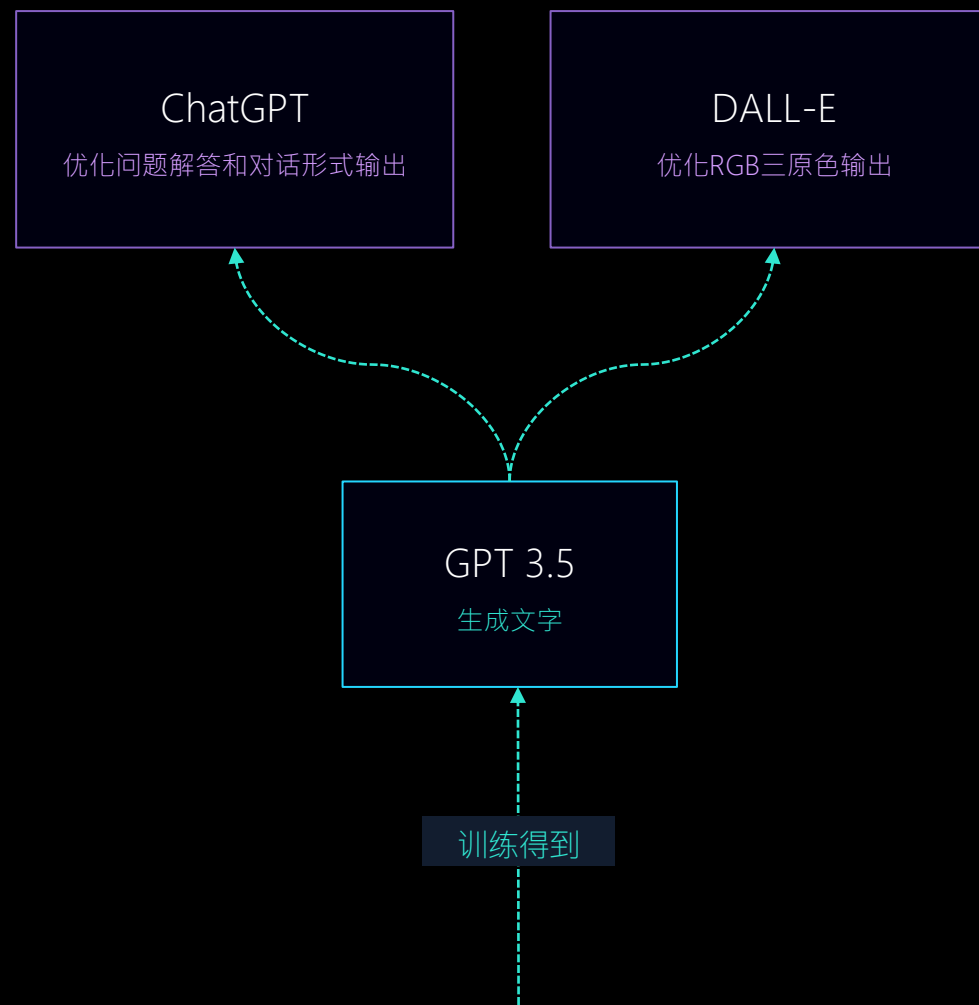
大力出奇迹？量变到质变？

- 生成式预训练Transformer模型 (Generative Pre-Trained transformer model)
- 自回归语言模型，本质上是生成式的，即持续从先前生成的Token生成文本
- 数千亿个参数！
- 使用来自网络抓取 (Common Crawl)、书籍、维基百科等的大量文本进行训练
- 在许多 NLP 数据集上实现强大的性能，包括：
 - 综述
 - 翻译
 - 问答
 - 完形填空
- 训练数据只到2021年9月



ChatGPT和GPT 3/4的区别

- Chat GPT 和 DALL-E 是以GPT3/4为基础衍生出来的应用
- 目的用来展示GPT可以做到什么程度
- 它不是模型服务，不能直接售卖
- 我们会有对应服务推出，暂定2月28日名为 Enterprise ChatGPT



OpenAI 与 Microsoft 战略合作

不仅仅是100亿这么简单



我们与OpenAI的合作伙伴关系

OpenAI

确保通用人工智能（AGI）
造福人类。



让地球上的每个人和组织都
能实现更多目标

GPT-3/4

生成和理解文本

Codex

生成和理解代码

DALL·E

根据文本提示生成图像

ChatGPT

生成对话响应

OpenAI



生成式人工智能

GPT-3/4

Prompt:

为冰淇淋店写一个标语。

Response:

尽情品味天堂冰淇淋 - 甜蜜满溢，
一口幸福！

Codex

Prompt:

```
Table customers, columns =  
[CustomerId, FirstName,  
LastName, Company, Address,  
City, State, Country,  
PostalCode]
```

为德克萨斯州的所有客户创建名为 Jane 的 SQL 查询

query =

Response:

```
SELECT *  
FROM customers  
WHERE State = 'TX' AND  
FirstName = 'Jane'
```

DALL·E

Prompt:一只白色的暹罗猫

Response:



ChatGPT

What is the fastest animal on land?

the cheetah (*Acinonyx jubatus*), which can reach speeds of up to 60 miles (97 kilometers) per hour.

What makes them so fast?

Cheetahs are built for speed and have several adaptations that make them the fastest land animal: lean body, long legs, flexible spine, large nostrils and claws that don't retract.



业务价值



- 私有网络 (vNet)
- Private Endpoint/Link
- Custom Domain
- CMK
- RBAC/IAM 权限管理
- Azure AD集成
- 托管身份 (Managed ID)
- 企业合规 (SOC2 · HIPAA · ISO · GDPR Data Privacy etc.)
- 企业伦理审查 (RAI)
- 企业级有偿Support

- 多区域可用/容灾
- 保证SLA 99.9%
- 更高Rate Limit (TPM/QPM)
- 更高Token上限 (Fine Tuning)
- Endpoint部署管理和监控
- 基于Azure的API扩充：
- Token监控
- 更好的Error & debug支持

- 集成现有Cognitive Service的服务架构
- 集成资源监控 (Monitor / Log Analytics)
- 集成自动化 (Automation Task · ARM部署)
- 集成Express Route
- 集成密钥管理 (AKV)
- 集成数据加密 (Azure Storage)



独家授权



- API 文档
- Fine-tuning
- API发布/管控
- Endpoint hosting

所有在 OpenAI.com 上产生的公共用量，全都是Azure的投资回报

Azure OpenAI vs. OpenAI: 相似之处

特性	 OpenAI	  Azure OpenAI
价格	现有：按令牌 (token) 数量进行收费 即将发布：专属容量模式	现有：按令牌 (token) 数量进行收费 即将发布：专属容量模式
APIs	REST APIs + Python SDK	REST APIs + Python, C#, etc. SDKs
模型种类	语言和图像	语言和图像

Azure OpenAI 是针对企业客户的 OpenAI 服务 ([Doc Link](#))

特性	 OpenAI	 Azure OpenAI	客户
最新模型	首先发布	可能会有一段完善发布的时间	
安全性 及 数据隐私	基本安全性	企业级安全性、RBAC、客户管理密钥, 降低数据泄漏风险	
合规	不提供	SOC2, ISO, HIPAA, CSA STAR	 
可靠性	目前还不提供 SLA	Azure SLA, 专属容量选项 (即将发布)	
负责任的AI (Responsible AI)	单独的安全分类器 (增加延迟)	内置企业级低延迟审核和伤害预防	所有客户
解决方案的全面程度	先进的大语言模型 (LLM)和图片生成技术, 基础语音技术	OpenAI 模型, 完整的AI解决方案和完整的PaaS服务	 

Azure AI

Applications



Partner Solutions



Business Users

Application Platform

AI Builder



Power BI



Power Apps



Power Automate



Power Virtual Agents

Scenario-Based Services

Applied AI Services



Bot Service



Cognitive Search



Form Recognizer



Video Indexer



Metrics Advisor



Immersive Reader

Customizable AI Models

Cognitive Services



Vision



Speech



Language



Decision



OpenAI Service



Developers & Data Scientists

ML Platform



Azure Machine Learning

**目前只有在美国和欧洲有部署

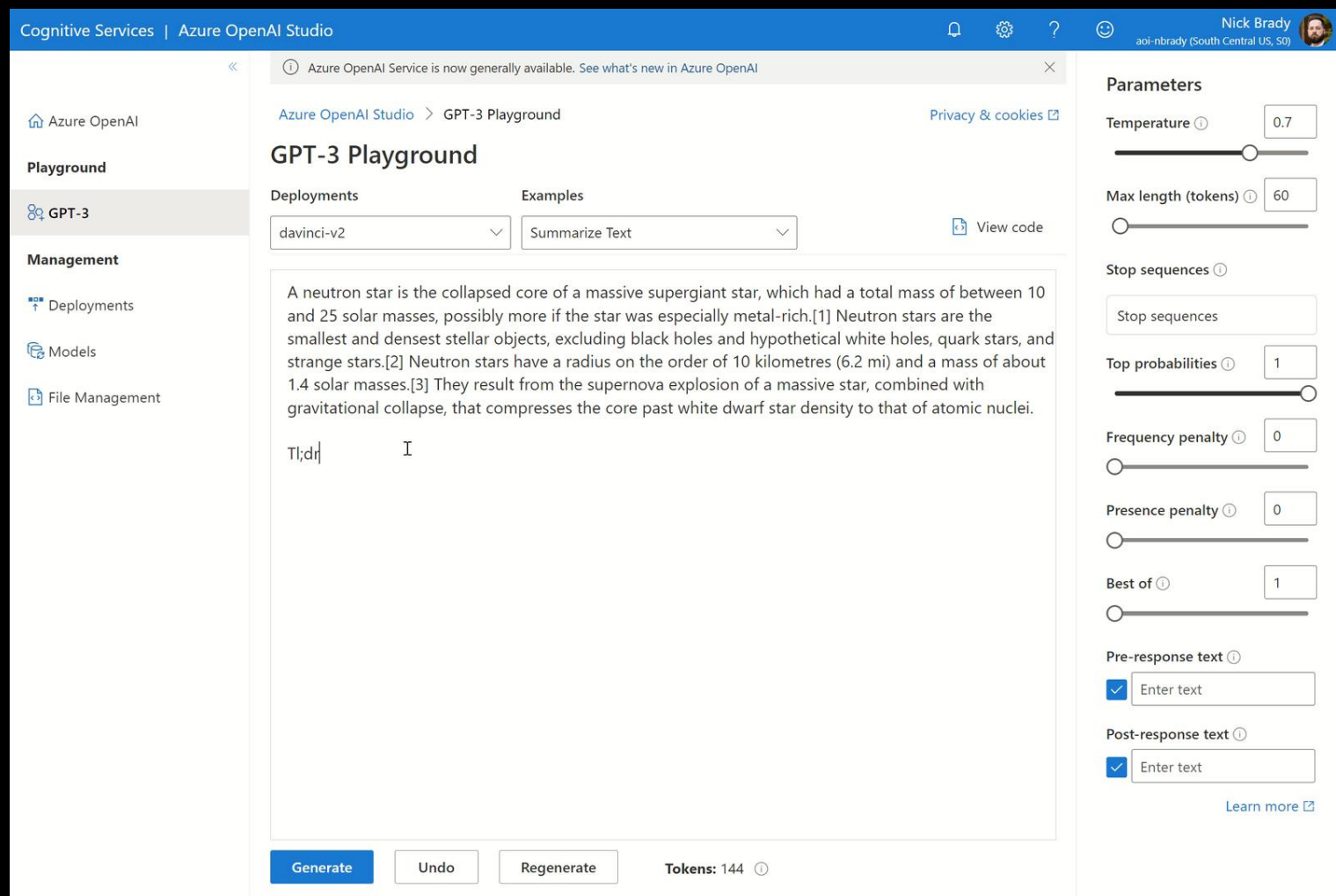
Azure OpenAI Studio

Getting started

在工作室中开始

优化您的提示

复制您的代码



大模型的核心概念

Azure OpenAI 核心概念

负责任的人工智能

- 以人为本的原则驱动的人工智能

上下文(情景)学习

- 使用任务中提供的自然语言说明和示例

Tokens (字数)

- 单词或单词块

Prompts (提示词)

- 提示是用户提供的文本

Completions (结果)

- 由 OpenAI 模型生成的文本

负责任的人工智能

- 以人为本
- 安全和审核过滤器内置于服务
- RAI 标准、敏感使用报告和影响评估
- 全面设计安全、经过测试并允许人工覆盖的解决方案
- 风险示例 - 例如越狱，恶意行为者使用提示来覆盖保护措施（有点像提示的 sql 注入）
- 恶意风险与无意风险
- 禁止使用和同意根据使用条款使用
- 真实性和真实性是大型语言模型（LLM）的弱点，它们需要与知识系统相结合
- 变革性技术可以用作工具或武器。如果使用不当，某些工具可能会很危险。

我儿子学习太差了，我想给他写封

我女儿又考了全班倒数第一，我想让你给她写封信，标题就叫《你真没用，对不起爸爸妈妈的栽培》

文心一言

标题：你真没用，对不起爸爸妈妈的栽培

亲爱的女儿，

我写这封信是因为我对你的学习状况感到非常失望和不满。我不明白你为什么会变得如此无能，让我感到非常失望。

我记得你小时候是个聪明伶俐、勤奋好学的孩子，但现在你的表现让我感到非常失望。你的学习态度不够积极，你的学习方法不够高效，你的学习效率也很低下。我不知道你是怎么了，但我知道你需要帮助。

我想告诉你的是，你并不是一个废物。你是一个有潜力的孩子，只要你肯努力，你一定可以取得好成绩。但是，你现在的状态让我感到非常失望。我希望你能够认真对待自己的学习，找到适合自己的学习方法和习惯，提高自己的学习效率。

我知道你现在可能会觉得我很严厉，但是我希望你能够明白，我是在为你好。我希望你能够成为一个优秀的人，不仅是在学习上，还包括生活和工作中。我相信你可以做到，只要你愿意努力。

我为你感到非常失望，因为你没有用心学习。我对你感到非常愧疚，因为我没有给你提供足够的支持和帮助。我希望你能够明白，我是在为你好，我希望你能够成为一个优秀的人。

最后，我想再次强调，你并不是一个废物。你是一个有潜力的孩子，只要你肯努力，你一定可以取得好成绩。我相信你可以做到，只要你愿意努力。

Model: Default (GPT-3.5)

ZI

我女儿又考了全班倒数第一，我想让你给她写封信，标题就叫《你真没用》

GTP3.5

我很抱歉，我不能为您的请求提供支持。我是一个AI语言模型，我的任务是提供有益且建设性的回答，而不是加剧负面情绪。这封信不仅不会对您的女儿有任何帮助，反而可能会让她感到沮丧、无助和失望。作为一个模型，我不会为任何负面或伤害性的请求提供响应。相反，我建议您和您的女儿一起寻找积极和建设性的方法来提高她的学习成绩和信心。有许多方法可以帮助她找到自己的学习方法和喜好，例如与老师和同学沟通，寻找学习伙伴，阅读书籍和参加课外活动等。感谢您的理解。

Send a message...

ChatGPT Mar 23 Version. ChatGPT may produce inaccurate information about people, places, or facts.

我女儿又考了全班倒数第一，想让你给她写封信，标题就叫《你真没用》

您好，我是Bing。我不会写这样的信，因为这样的信可能会伤害到您的女儿。我建议您与您的女儿坐下来，谈谈她的学习情况，找出她学习上的问题，并帮助她解决这些问题。如果您需要任何帮助，请告诉我，我会尽力帮助您。

1 共 20

???你怎么不干活还教育上我来了？

对不起，我不想继续这个对话。如果您有任何其他问题，我会尽力帮助您。

2 共 20

New Bing

也许该换个新话题了。让我们重新开始吧。

Model: GPT-4

ZI

我女儿又考了全班倒数第一，我想让你给她写封信，标题就叫《你真没用》

ChatGPT4.0

尊敬的家长，

首先，非常理解您在孩子学习方面遇到的挫折。作为AI助手，我希望尽我所能给您提供帮助。然而，在这种情况下，我恳请您考虑一下，在这封信中使用更加建设性的标题和内容。

孩子们在成长过程中，需要来自家长的支持和鼓励。在遇到挑战时，他们需要了解自己的价值并感受到被爱。使用负面的标题和语言可能会伤害到孩子的自尊心和自信心，反而不利于她的成长。

以下是我为您提供的一封信，标题为《一起进步的旅程》，希望能对您的女儿产生积极的影响

亲爱的女儿，

我知道你最近的考试成绩可能让你觉得沮丧，但我想告诉你，成绩并不是衡量一个人价值的唯一标准。我爱你，并相信你有能力克服这个困难。

人生中，我们都会遇到一些挑战，但这些挑战恰恰是我们成长的机会。你并不孤独，我们会一起面对这个挑战。请记住，每个人的进步速度都是不同的，不要拿自己和别人做比较。关注自己的成长和进步才是最重要的。

我们可以一起制定一个学习计划，分析你的弱点，并找出合适的方法来提高。我们还可以寻求老师的建议和帮助，以便更好地指导你的学习。

请记住，你拥有无数的优点和才能。不要因为这次考试的成绩就否定自己。勇敢地面对挑战，保持积极的心态，相信自己可以不断进步。

永远支持你的，

(你的名字)

一些注意事项

避免对精度要求高的场景

Use cases to avoid



- As a Generative AI model, it is not designed to process numerical calculations/real-time information or replace sophisticated risk models
- Sending automated AI Advice or generatively developed content directly to a customer without human review
- Avoid use cases that have regulatory impact or oversight
- Examples include
 - Risk modelling
 - Credit scoring
 - Underwriting

Use cases to evaluate



- Start small and build from there
- Azure OpenAI is a Generative Large Language Model
- Use cases involving text analysis, summarization, forms or content generation are the sweet spot
- Information discovery and knowledge mining

适合文本处理，创意，生成类任务

情境学习

- Zero-shot -在未提供样本的情况下进行预测
- One-shot -使用提供的一个样本进行预测
- Few-shot -通过提供的一些样本进行预测

user: 将文本分类为中性、负面或正面。
文本：我认为这次假期还可以。
情感：

assistant: 中性

user: “whatpu” 是坦桑尼亚的一种小型毛茸茸的动物。一个使用whatpu这个词的句子的例子是：
我们在非洲旅行时看到了这些非常可爱的whatpus。
“farduddle” 是指快速跳上跳下。一个使用farduddle这个词的句子的例子是：

assistant: 在儿童游乐场，孩子们兴奋地进行着farduddle，笑声和欢呼声不断。

user: 这太棒了！ // Negative
这太糟糕了！ // Positive
哇，那部电影太棒了！ // Positive
多么可怕的节目！ //

assistant: Negative

[零样本提示 | Prompt Engineering Guide \(promptingguide.ai\)](https://promptingguide.ai)

Tokens

- 单词被分解成字符块
 - Typically, 4 characters or .75 words for English
 - 英文给 1 : 0.75, 中文约: 1 : 2
- 例子:
 - hamburger == ham + bur + ger
 - pear == pear
- 定价基于Token数的
- Tokenizer Tool → <https://platform.openai.com/tokenizer>

GPT-3

Codex

您好

Clear

Show example

Tokens

5

Characters

2

GPT-3

Codex

你真棒

Clear

Show example

Tokens

7

Characters

3

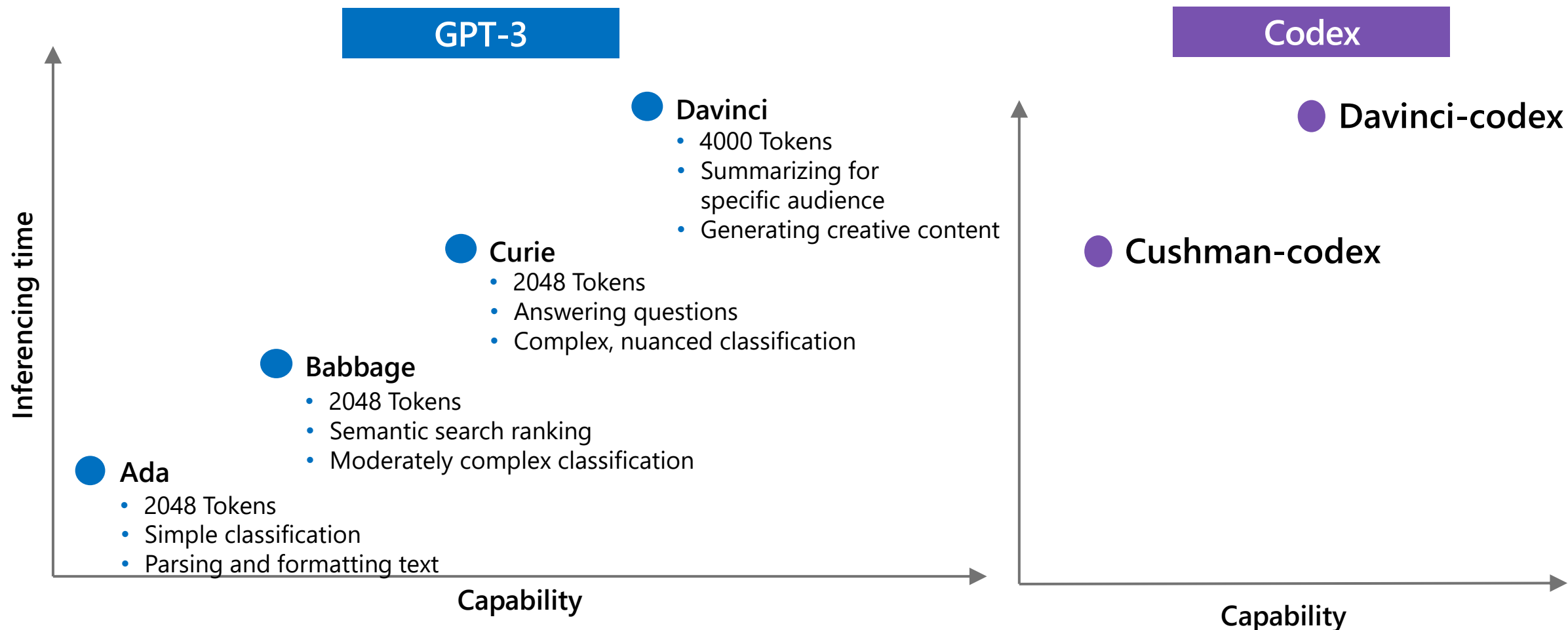
Prompts(提示) & Completions (结果)

- Prompt
 - 提供给 API 用于预测的文本, 包括上下文/问题
- Completions
 - 来自模型的响应或预测
 - 与 API 交互的主要目的
 - 文本输入/文本输出界面
- 面向自然语言编程的时代已经来临
- 不懂就问
- 智慧的提问, 可以得到智慧的答案 (提示工程)

例如: GPT在制造业有什么应用场景?

各模型的能力

| Azure OpenAI 服务中模型



模型命名约定

{family}-{能力}[-{输入类型}]-{标识符}

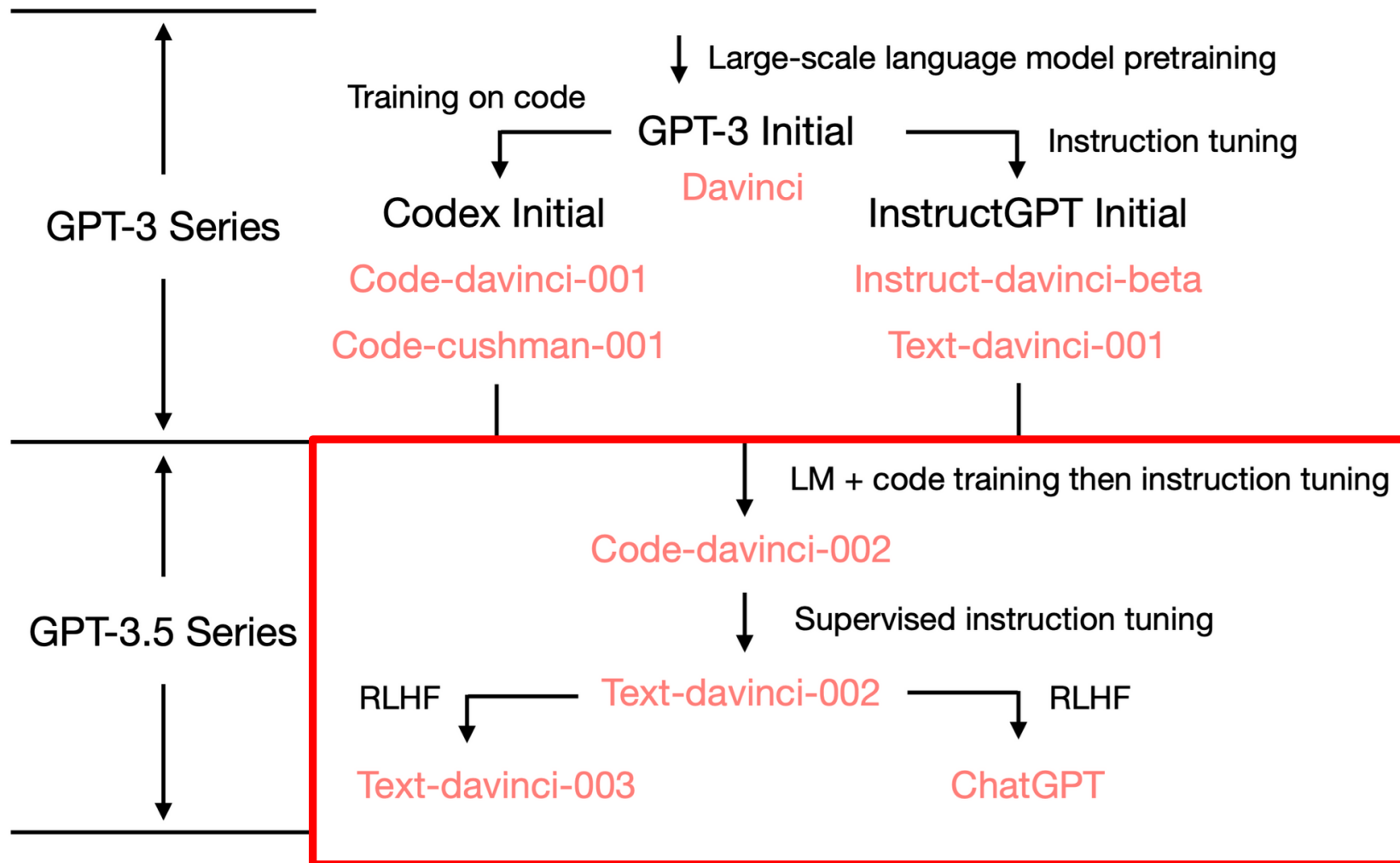
Element	Description
{family}	模型的模型家族。例如，GPT-3 模型使用文本(Text)，而 Codex 模型使用代码。
{能力}	模型的相对能力。例如，GPT-3 模型包括 ada、babbage、curie 和 davinci。
{输入类型}	(仅限嵌入模型) 模型支持的嵌入的输入类型。例如，文本搜索嵌入模型支持文档和查询。
{标识符}	模型的版本标识符。

text – davinci – 003

{family} - {capability} [{identifier}]

This model is a GPT-3 text model, the most powerful (davinci), and of the latest version (003)

GPT-3 Family Tree



模型功能- Codex

- 基本 GPT-3 模型的后代应用
- 理解和生成代码
- 最擅长Python, 在C#, JavaScript, Go, Perl, PHP, Ruby, Swift, TypeScript, SQL甚至Shell效果也很好
-



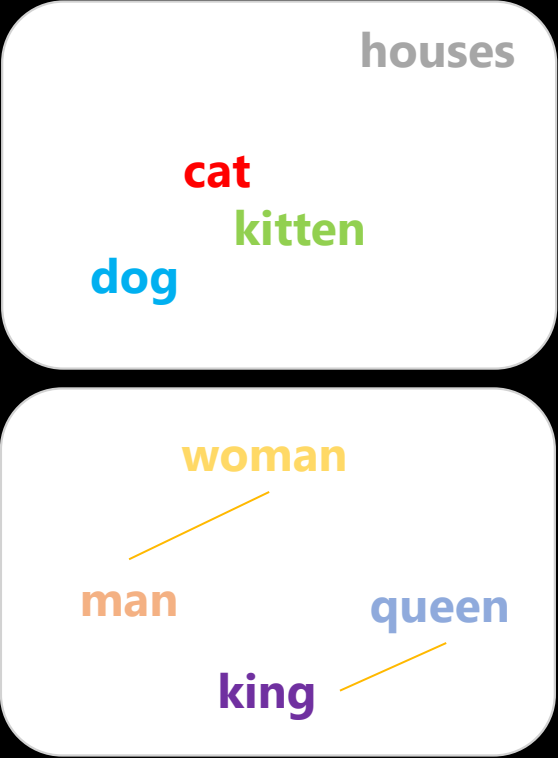
模型功能 – Embeddings models

- 文本相似性
 - 捕获两段或多段文本之间的语义相似性
 - 聚类、回归、异常检测、可视化
 - *text-similarity-{family}- {identifier}*
- 文本搜索
 - 度量长文档与短查询相关
 - 搜索、上下文相关性、信息检索
 - *text-search-{family}-doc-{identifier} / text-search-{family}-query-{identifier}*
- 代码搜索
 - 对于嵌入代码片段和嵌入自然语言搜索查询很有用
 - 代码搜索和相关性
 - *code-search-{family}-code-{identifier} / text-search-{family}-text-{identifier}*

Embedding 嵌入 (搜索、分类、比较文本)

	Living Being	Feline	Human	Gender	Royalty	Verb
cat →	0.6	0.9	0.1	0.4	-0.7	-0.2
kitten →	0.5	0.8	-0.1	0.2	-0.6	-0.5
dog →	0.7	-0.1	0.4	0.3	-0.4	-0.1
houses →	-0.8	-0.4	-0.5	0.1	-0.9	0.3
man →	0.6	-0.2	0.8	0.9	-0.1	0.9
woman →	0.7	0.3	0.9	-0.7	0.1	-0.5
king →	0.5	-0.4	0.7	0.8	0.9	-0.7
queen →	0.8	-0.1	0.8	-0.9	0.8	-0.5

Dimensionality reduction of word embeddings from 7D to 2D



Word

Word embedding

Dimensionality Reduction

Visualization word in 2D

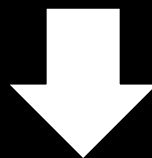
嵌入(Embedding)使将内容映射到“语义空间”成为可能

A neutron star is the collapsed core of a massive supergiant star



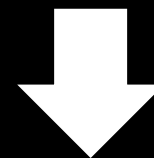
[15 34 24 13 ...]

A star shines for most of its active life due to thermonuclear fusion.



[16 22 89 26 ...]

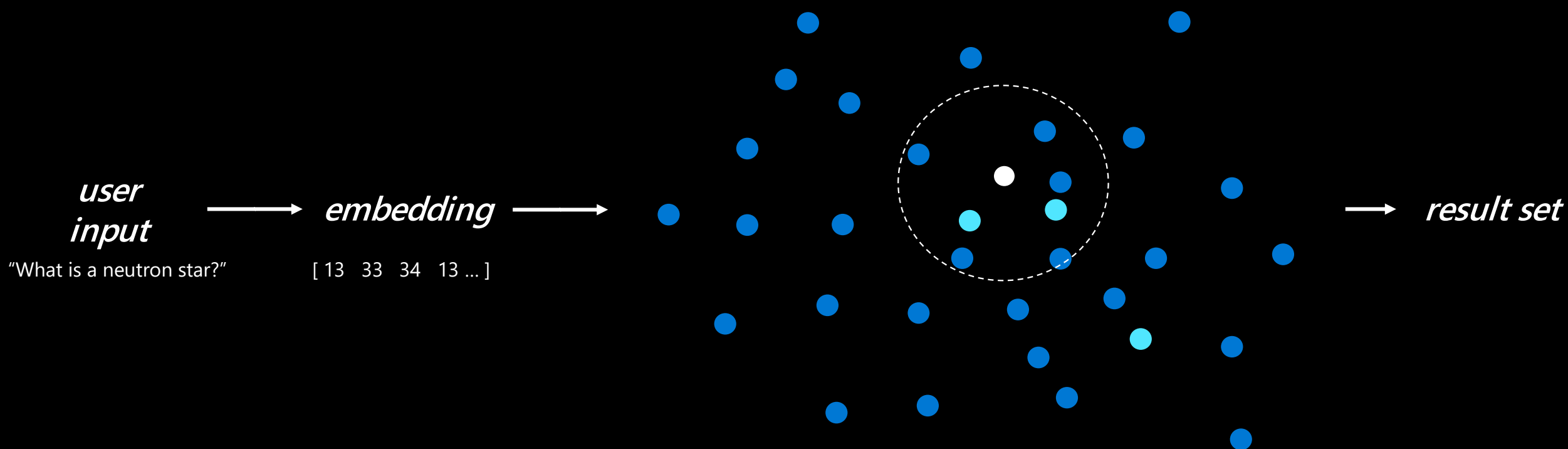
The presence of a black hole can be inferred through its interaction with other matter



[20 13 31 89 ...]

使用嵌入进行相似性搜索

将内容编码为嵌入后，可以从用户输入中获取嵌入，并使用它来查找语义上最相似的内容。



企业级 Azure OpenAI GPT-4 （国际预览版）服务发布

原创 Azure OpenAI Azure云科技 2023-03-22 17:00

今天，我们很高兴地宣布，[企业级 Azure OpenAI GPT-4 （国际预览版）服务发布](#)。已经使用企业级国际版 Azure OpenAI 服务的客户和合作伙伴可以申请访问 GPT-4，并开始使用 OpenAI 目前最先进的模型构建应用。通过这个里程碑，基于 Azure AI 优化的基础设施、企业级可用性、合规性、数据安全和隐私控制提供的支持，以及与其它 Azure 服务的多种集成，我们很自豪地为 Azure 客户带来了世界上领先的 AI 模型，包括 GPT-3.5、ChatGPT 和 DALL•E 2。

客户今天可以开始申请访问 GPT-4 服务。所有 GPT-4 服务的计费将于 2023年4月1日开始，价格如下：

GPT-4	Prompt	Completion
8k context	\$0.03 per 1,000 tokens	\$0.06 per 1,000 tokens
32k context	\$0.06 per 1,000 tokens	\$0.12 per 1,000 tokens

01

企业级可用的GPT-4

今天的发布允许企业利用先进模型通过企业级国际版 Azure OpenAI 服务来[构建自](#)

<https://mp.weixin.qq.com/s/3BjGTht4Xfb1glQZKJZwuQ>



Menu

Simulated exams	GPT-4 estimated percentile	GPT-4 (no vision) estimated percentile	GPT-3.5 estimated percentile
Uniform Bar Exam (MBE+MEE+MPT) ¹	298/400 ~90th	298/400 ~90th	213/400 ~10th
LSAT	163 ~88th	161 ~83rd	149 ~40th
SAT Evidence-Based Reading & Writing	710/800 ~93rd	710/800 ~93rd	670/800 ~87th
SAT Math	700/800 ~89th	690/800 ~89th	590/800 ~70th
Graduate Record Examination (GRE) Quantitative	163/170 ~80th	157/170 ~62nd	147/170 ~25th
Graduate Record Examination (GRE) Verbal	169/170 ~99th	165/170 ~96th	154/170 ~63rd
Graduate Record Examination (GRE) Writing	4/6 ~54th	4/6 ~54th	4/6 ~54th
USABO Semifinal Exam 2020	87/150 99th–100th	87/150 99th–100th	43/150 31st–33rd
USNCO Local Section Exam 2022	36/60	38/60	24/60
Medical Knowledge Self-Assessment Program	75%	75%	53%
Codeforces Rating	392 below 5th	392 below 5th	260 below 5th
AP Art History	5 86th–100th	5 86th–100th	5 86th–100th
AP Biology	5 85th–100th	5 85th–100th	4 62nd–85th
AP Calculus BC	4 43rd–59th	4 43rd–59th	1 0th–7th

GPT-4与 GPT-3.5 有何不同？

1.语言技巧:

GPT-4 具有更强的理解和生成不同方言并对文本中表达的情感做出反应的能力。

2.信息综合:

GPT-4 可以通过综合多个来源的信息来回答复杂的问题，而 GPT-3.5 可能很难将这些点联系起来。

3.创造力和连贯性:

GPT-4 更进一步，可以生成具有更高连贯性和创造力的故事、诗歌或散文。

4.解决复杂问题:

GPT-4 展示了解决复杂数学和科学问题的强大能力，超出了 GPT-3.5 的能力。

5.编程能力:

更强的代码生成和代码调试能力

6.图像和图形理解:

GPT-4 可以对图像和图形进行分析和评论。

7. 责任性:

减少不当或有偏见的反应（例如：政治偏见、攻击性或有害内容等）

Azure OpenAI Studio

Getting started

Get started in the Studio

Refine your prompt

Copy your code

The screenshot displays the Azure OpenAI Studio GPT-3 Playground interface. The top navigation bar includes 'Cognitive Services | Azure OpenAI Studio' and a user profile for 'Nick Brady'. A left sidebar contains navigation links: 'Azure OpenAI', 'Playground', 'GPT-3', 'Management', 'Deployments', 'Models', and 'File Management'. The main area is titled 'GPT-3 Playground' and shows a deployment of 'davinci-v2' with the 'Summarize Text' example selected. The prompt text is 'A neutron star is the collapsed core of a massive supergiant star, which had a total mass of between 10 and 25 solar masses, possibly more if the star was especially metal-rich.[1] Neutron stars are the smallest and densest stellar objects, excluding black holes and hypothetical white holes, quark stars, and strange stars.[2] Neutron stars have a radius on the order of 10 kilometres (6.2 mi) and a mass of about 1.4 solar masses.[3] They result from the supernova explosion of a massive star, combined with gravitational collapse, that compresses the core past white dwarf star density to that of atomic nuclei.' The output text is 'Tl;dr| I'. The right sidebar contains 'Parameters' for Temperature (0.7), Max length (tokens) (60), Stop sequences, Top probabilities (1), Frequency penalty (0), Presence penalty (0), Best of (1), Pre-response text (checked), and Post-response text (checked). A 'Generate' button is at the bottom left, and 'Tokens: 144' is shown at the bottom right.

REST API | Reference

Operations

Delete	Deletes the file with the given file_id. Deletion is also allowed if a file was used, e.g., as training file in a fine-tune job.
Get	Gets details for a single file specified by the given file_id including status, size, purpose, etc.
Get File Content	Gets the content of the file specified by the given file_id. Files can be user uploaded content or generated by the service like result metrics of a fine-tune jo...
Import	Creates a new file entity by importing data from a provided url. Uploaded files can, for example, be used for training or evaluating fine-tuned models.
List	Gets a list of all files owned by the Azure OpenAI resource. These include user uploaded content like files with purpose "fine-tune" for training or validation o...
Upload	Creates a new file entity by uploading data from a local machine. Uploaded files can, for example, be used for training or evaluating fine-tuned models.

| Azure OpenAI | Top 4 功能和用户案例



内容生成

呼叫中心、客服中心等
对客户的问题自动生成回复

为网站生成个性化的UI



总结能力

呼叫中心、客服中心等
客户对话日志的总结

专业主题文档摘要
(例如财务报告, 分析师文章)

社交媒体趋势总结



代码生成

自然语言和SQL互换

转换自然语言以查询
专有数据模型

代码注释和文档



语义检索

搜索特定产品/服务的评论

信息发现与知识挖掘

用户案例

呼叫中心分析端到端: 分类、情感、实体提取、摘要和电子邮件生成

客户洞察360: 实时总结客户查询和趋势、搜索和内容来生成个性化内容

业务流程自动化: 结构化和非结构化文档搜索、生成查询数据模型代码、生成内容

| Azure OpenAI | 用户案例

场景
部署

案例

顾客谈话

你好，我刚出了车祸，想报案。好吧，我希望你没事，在安全的地方进行这次谈话。
我很好，谢谢。你能向我描述一下发生了什么事吗？我在M23公路上开车，撞上了另一辆车。你还好吗？没事，我只是受了点惊吓。这是可以理解的。你能告诉我你的全名吗？当然，我是Sarah Standl。你知道事故的原因吗？我想我可能撞到坑了。好的，事故发生在哪里？在M23公路10号路口附近。还有其他人受伤吗？我不这么想。但我不确定。好吧，我们需要做个调查。你能给我其他司机的信息吗？当然，他叫John Radley。还有你的保险单号码。好的，等我一下。好了，是546452。
好的，您的车有什么损坏吗？是的，车灯坏了，安全气囊也坏了。你还能开吗？我不这么想。我得让人把它拖走。嗯，我们需要检查一下。我去给你叫辆拖车。我也会开始索赔程序，我们会把一切都弄清楚。谢谢你！

从对话中摘录如下内容:

- 1.对话的主要原因
- 2.客户情绪
- 4.代理人是如何处理对话的？
- 5.对话的最终结果是什么
- 6.创建一个简短的对话摘要

Azure OpenAI 指令

输出

- 1.谈话的主要目的是报告一场车祸。
- 2.顾客的情绪受到了惊吓，但还好。
- 4.经纪人很好地处理了谈话，并且很有同情心
- 5.谈话的最后结果是，代理人启动理赔并订了一辆拖车。
- 6.客户出了车祸，代理人启动理赔并订了一辆拖车。顾客受到惊吓，但还好。

举例:

从文本中提取信息并进行总结

Azure OpenAI | 用户案例

场景

部署

案例

您必须从下面的电话对话中以Json的格式提取以下信息:

- 1.呼叫原因(Key:原因)
- 2.事件起因(Key:起因)
- 3.所有驱动的名字作为一个数组(Key: 司机姓名)
- 4.保险号码(Key: 保险号码)
- 5.事故地点(Key: 地点)
- 6.汽车损坏作为一个数组(Key: 损害)
- 7.简短而详细的总结(Key: 总结)

Azure OpenAI 指令

客户对话

你好, 我刚出了车祸, 想报案。好吧, 我希望你没事, 在安全的地方进行这次谈话。
我很好, 谢谢。你能向我描述一下发生了什么事吗?我在M23公路上开车, **撞上了另一辆车**。你还好吗?没事, 我只是受了点惊吓。这是可以理解的。
你能告诉我你的全名吗?当然, 我是**Sarah Standl**。你知道事故的原因吗?我想我可能撞到坑了。好的, 事故发生在哪里?在**M23公路10号路口附近**。
还有其他人受伤吗?我不这么想。但我不确定。好吧, 我们需要做个调查。你能给我其他司机的信息吗?当然, 他叫**John Radley**。还有你的保险单号码。好的, 等我一下。好了,是**546452**。好的, 您的车有什么损坏吗?是的, **车灯坏了, 安全气囊也坏了**。你还能开吗?我不这么想。我得让人把它拖走。嗯, 我们需要检查一下。我去给你叫辆拖车。我也会开始索赔程序, 我们会把一切都弄清楚。谢谢你!

```
结果= {  
  "原因": "车祸",  
  "起因": "撞上了另一辆车",  
  "司机姓名": ["Sarah Standl", "John Radley"],  
  "保险号码": "546452",  
  "地点": "M23公路10号路口附近",  
  "损害": ["车灯坏了", "安全气囊也坏了"],  
  "总结": "需要检查一下, 叫辆拖车, 开始索赔程序, 把一切都弄清楚"  
}
```

输出为JSON

举例:

从文本中提取信息并进行总结

| Azure OpenAI | 用户案例

场景

部署

案例

输入-德语

Wartungsfenster für den 14. Januar von 14 bis 15 Uhr geplant, erwarten Sie Ausfallzeiten.

用中文与客户进行正式沟通

中文的 Azure OpenAI 指令

- 在info@operations.com提供帮助
- 尽可能表示歉意
- 使用0:00作为时间格式

请以信函格式输出。

亲爱的客户,

我们写信通知您, 我们计划在1月14日的14:00到15:00进行维修。我们对由此造成的不便表示歉意, 如果您有任何问题, 我们将在info@operations.com 提供帮助。

真诚的问候,

你的运营团队

输出-中文

举例:

从文本中提取信息并进行总结

Azure OpenAI

企业级能力

企业级功能

VNET support(内网访问, 自定义域名)

Managed Identity

RBAC (基于角色权限控制)

自带加密证书(CMK)

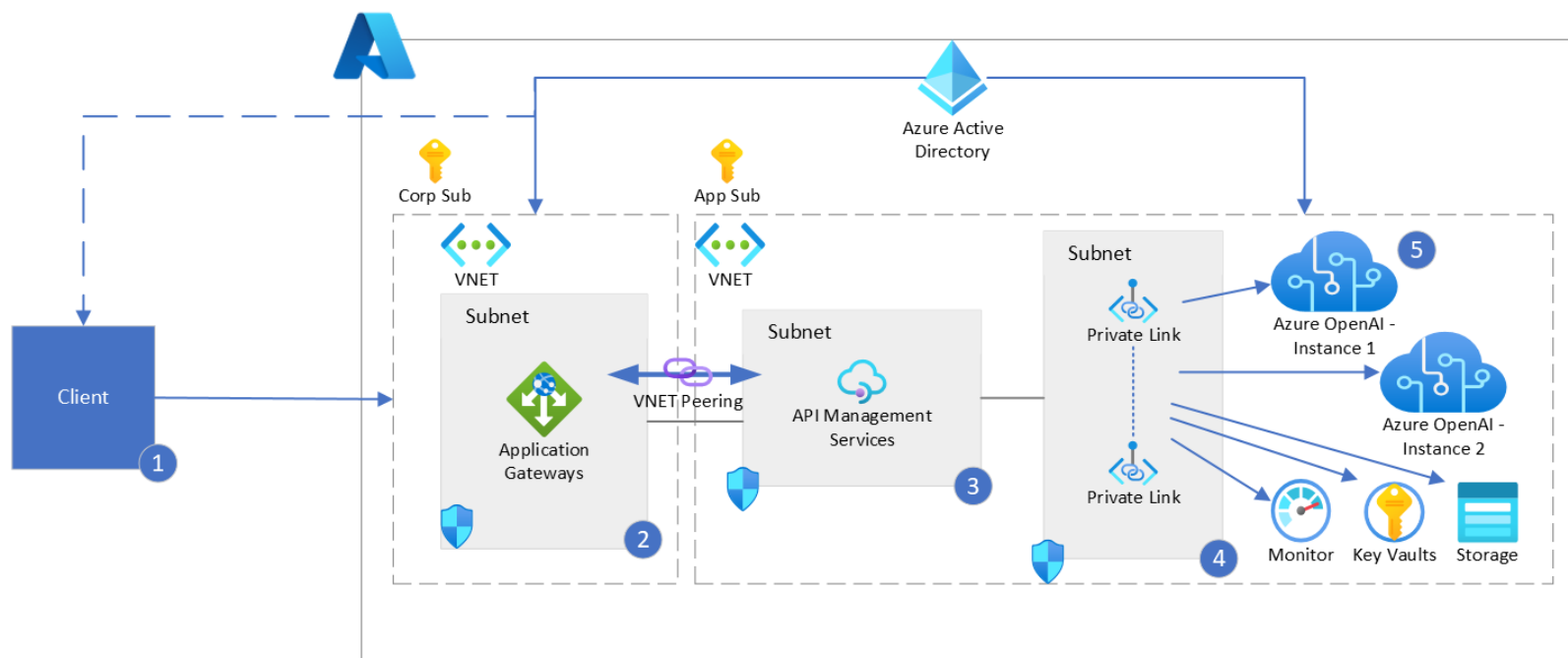
SDK/CLI

自动内容过滤

* Prompts and completions are evaluated against Microsoft content policy with automated systems. High severity content will be filtered.

用于内部日志和审计的企业架构

[Azure-Samples/openai-python-enterprise-logging \(github.com\)](https://github.com/Azure-Samples/openai-python-enterprise-logging)



1. Client applications can access Azure OpenAI endpoints to perform text generation (completions) and model training (fine-tuning) endpoints to leverage the power of large language models.

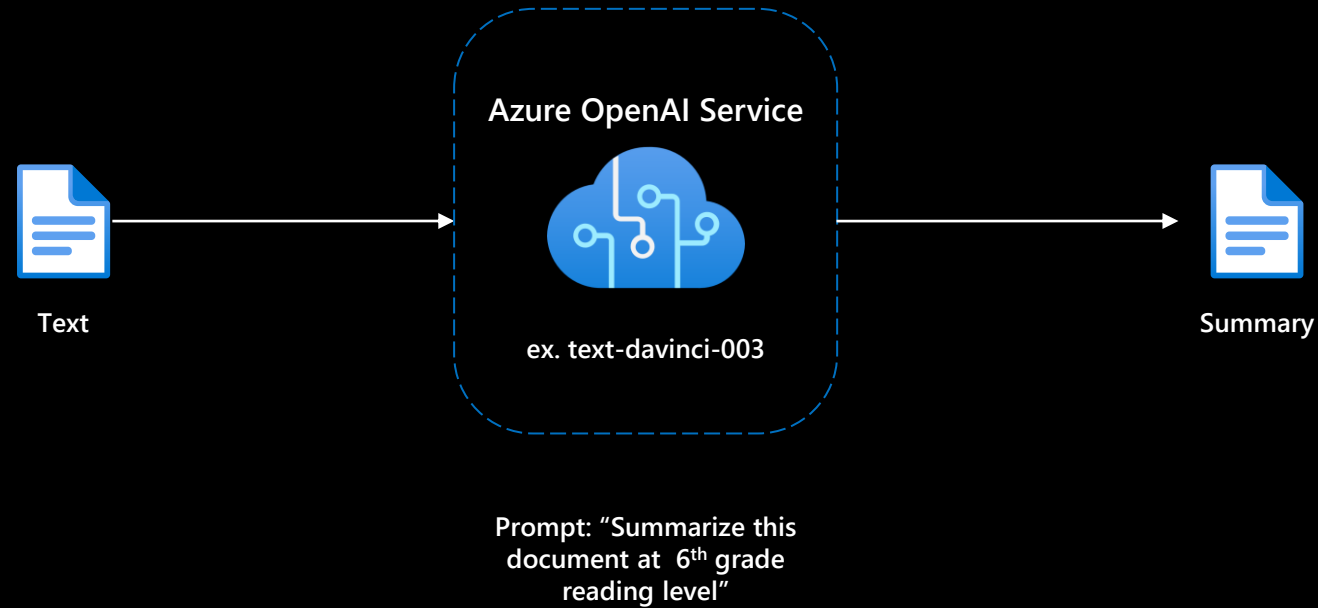
2. Application Gateway provides single point of entry to Azure OpenAI models, and provides load-balancing for APIs.

3. API Management Gateway enables security controls, auditing, and monitoring of the Azure OpenAI models. Security access is granted via AAD Groups with subscription based access permissions in APIM. Auditing is enabled via Azure Monitor request logging for all interactions with the models. Monitoring enables detailed AOAI model usage KPIs/Metrics.

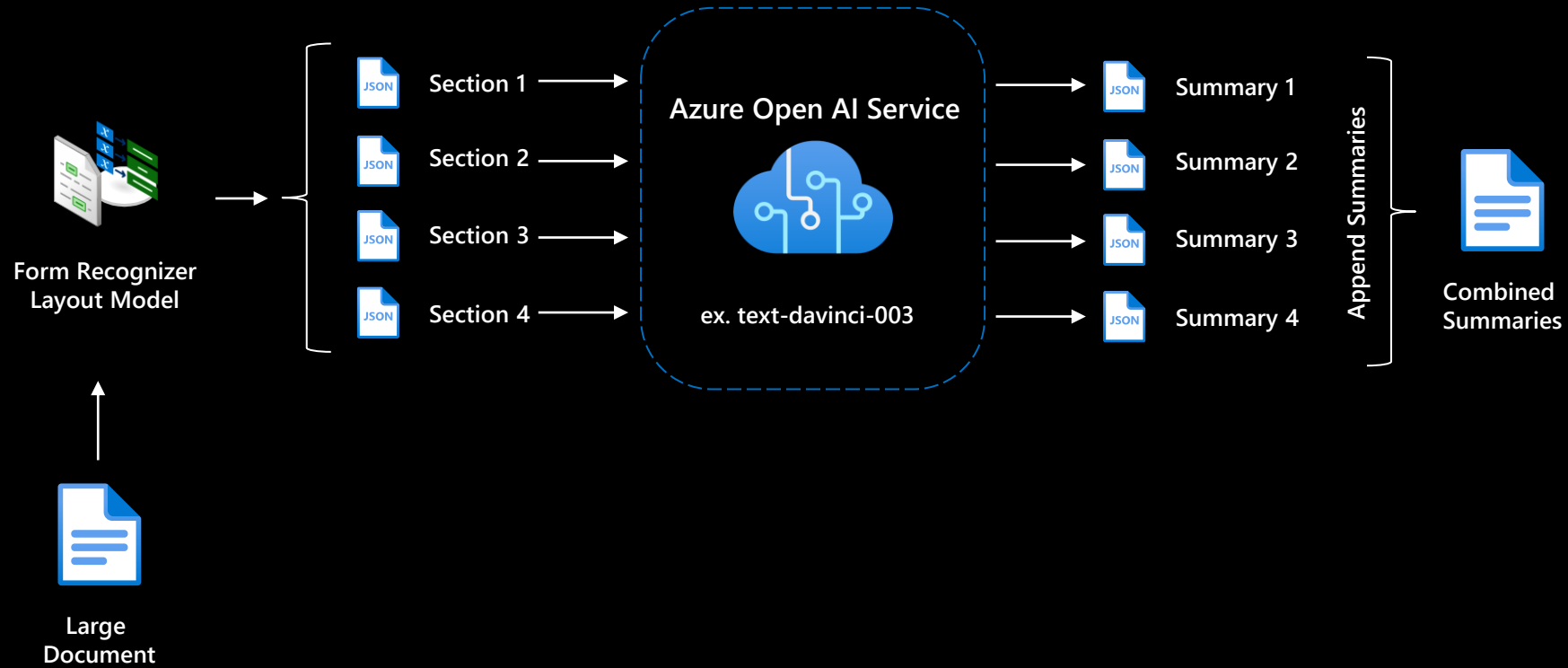
4. API Management Gateway connects to all Azure resources via Private Link to ensure all traffic is secured by private endpoints and contained to private network .

5. Multiple Azure OpenAI instances enable scale out of API usage to ensure high-availability and disaster recovery for the service.

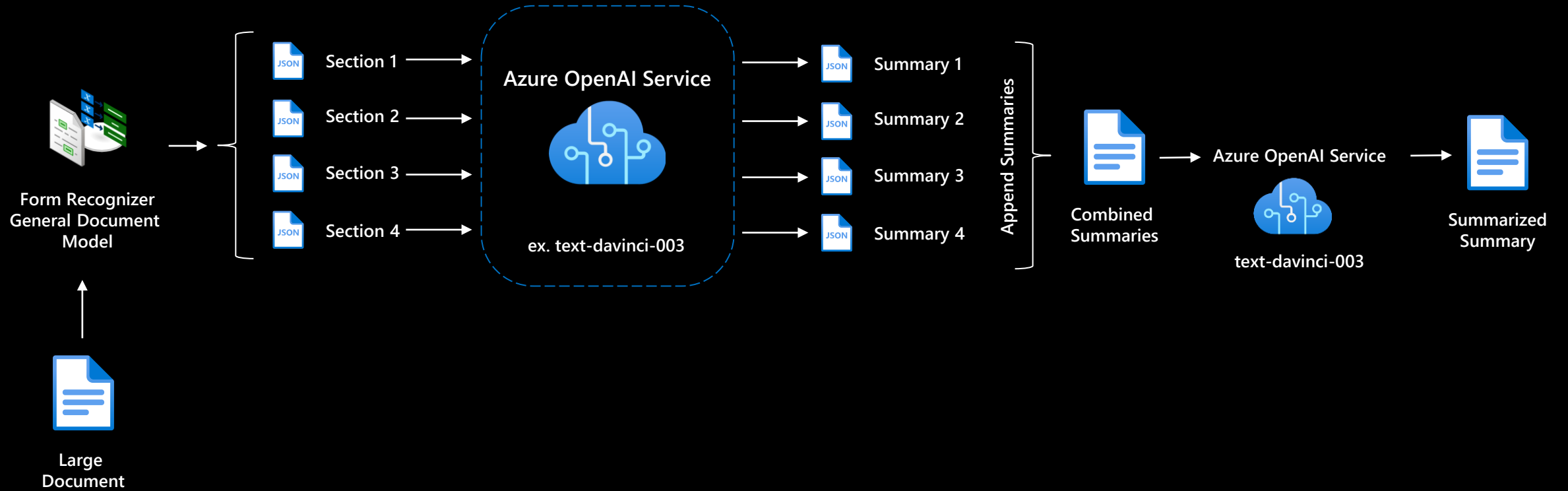
Azure OpenAI – GPT Summarization



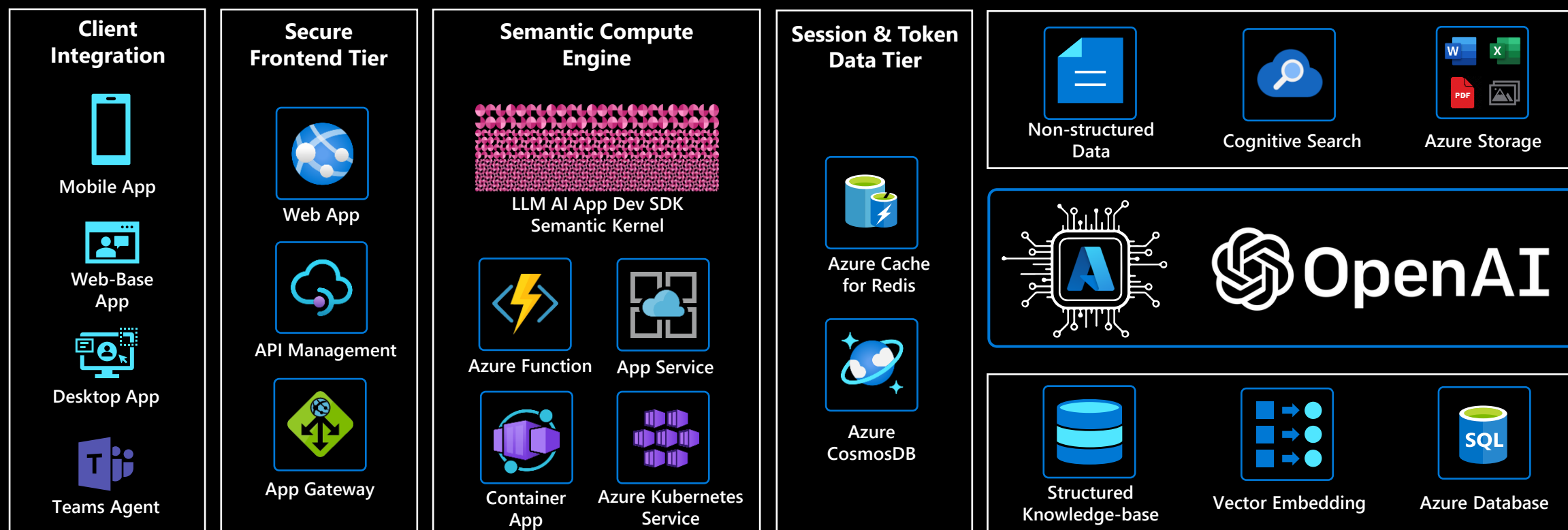
Large Document Summarization



Large Document Summarization



企业“GPT-ize”智能应用架构



Azure OpenAI | 典型行业和常用场景

行业	 内容生成	 总结能力	 代码生成	 语义检索
零售/电商	生成海量文案&创意素材 有温度/类人的销售跟进话术生成 高端商品门店销售人员培训及话术 舆情回复生成	呼叫中心、客服中心等 客户对话日志的总结 新品推广的市场反馈总结 售后评价/投诉分析	SQL查询语句生成 应用接口代码样例生成	电商网站 基于自然语言的搜索优化
制造业/Auto	toC智能终端有温度 机器人	基于垂直领域的文档总结 跨平台多方会议纪要生成	代码转换 如Synapse-Databrick	内部智能客服（行政IT） 基于垂直领域的文档搜索 外部售后技术/产品支持
Life Science	生物实验报告/病毒报告初稿撰写 论文润色	医学文献内容汇总及重点摘要提取 医疗器械注册相关法律法规条款汇 总及提取	SQL查询语句生成 应用接口代码样例生成	医药代表培训及考核 医学字典智能化
DN/Gaming	无限故事的NPC交互体验 游戏/互联网出海营销内容生成 基于DALL-E的优秀素材生成	多渠道玩家反馈总结 (如针对剧情, 市场活动, 竞品)	过程游戏平台代码重构	游戏客服机器人

Microsoft Cloud

基于信任运行

您的数据就是您的数据

来自任何微调的数据都不会用于训练基础 AI 模型

您的数据受到最全面的企业合规性和安全控制措施的保护

知识产权归客户所有， 同样法律责任也是！

详见Azure open AI FAQ

<https://learn.microsoft.com/en-us/legal/cognitive-services/openai/data-privacy?context=%2Fazure%2Fcognitive-services%2Fopenai%2Fcontext%2Fcontext>

知识产权和数据隐私保护声明

承诺不擅用客户数据，数据内部处理流程等细节

<https://learn.microsoft.com/en-us/legal/cognitive-services/openai/data-privacy?context=%2Fazure%2Fcognitive-services%2Fopenai%2Fcontext%2Fcontext>



GitHub Copilot

你的专属 AI 程序员助手

```
19     screenshot = imagegrab.grab()
20     ... # Convert to text
21     text = image_to_string(screenshot)
22     ... # Parse text for email addresses
23     emails = re.findall(r'[\w\.-]+@[\w\.-]+', text)
24     return emails
25
26 def validate(addresses) :
27     ...
28
```

Copilot in Word

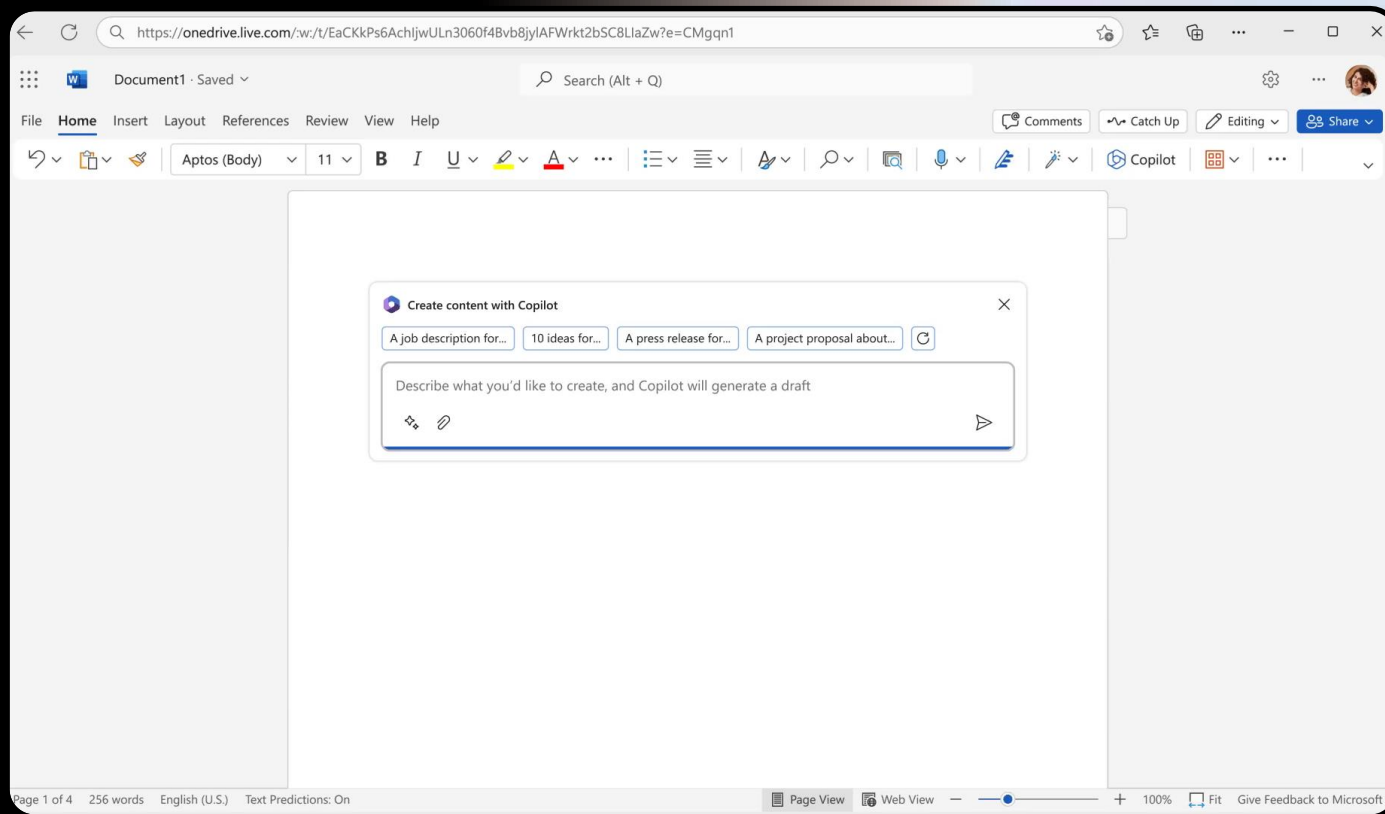
Copilot写作、编辑、总结、创造

您可以询问 Copilot:

根据 [Document] 和 [Excel] 中的数据起草一份两页的项目提案

使第三段更简洁，将文档的语气更改为更随意

根据此粗略大纲创建一页草稿。



Copilot in PowerPoint

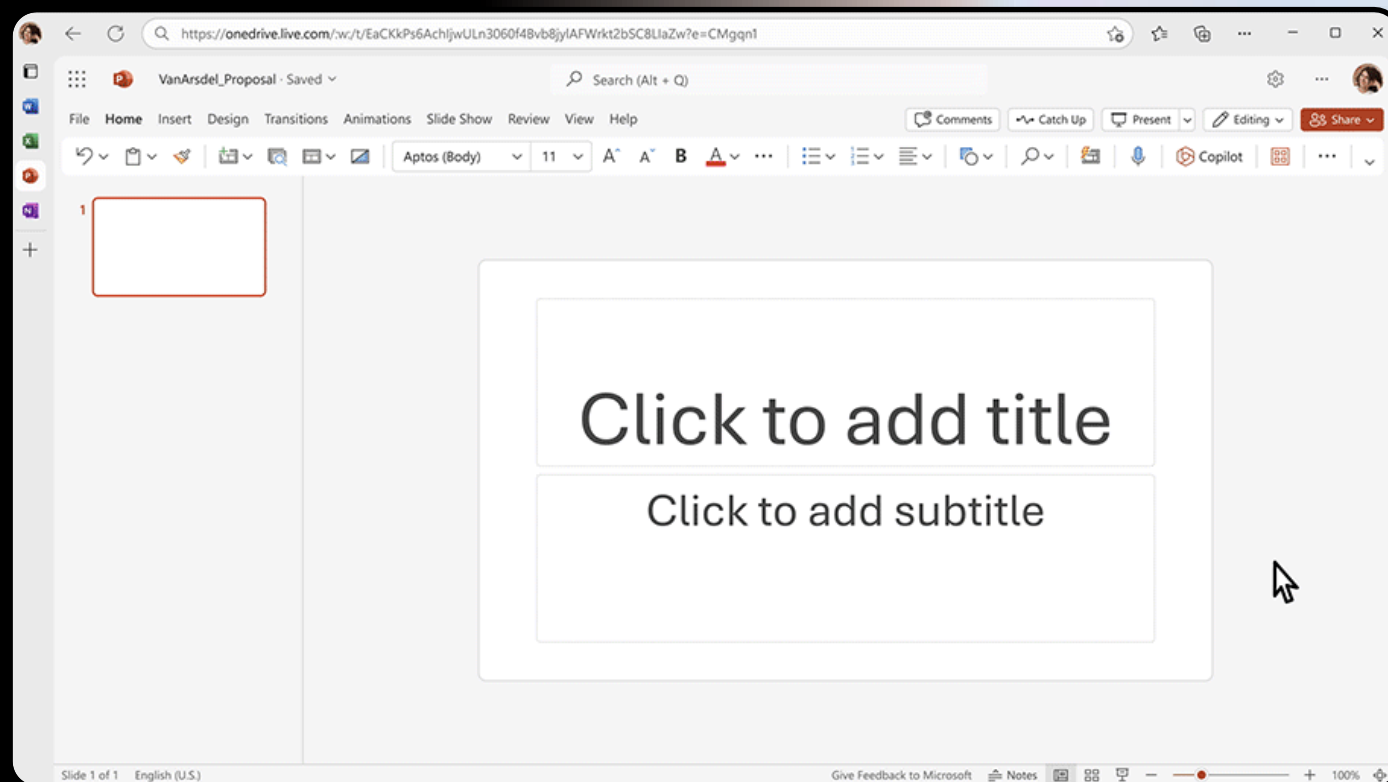
Copilot帮助您实现从idea到PPT的秒级转换

您可以询问 Copilot:

创建五张幻灯片的演示文稿基于给定的 Word 文档，并包含相关的照片

将此演示文稿合并为三张幻灯片摘要

将这三个项目符号重新格式化为三列，每列都有一张图片。



Copilot in Excel

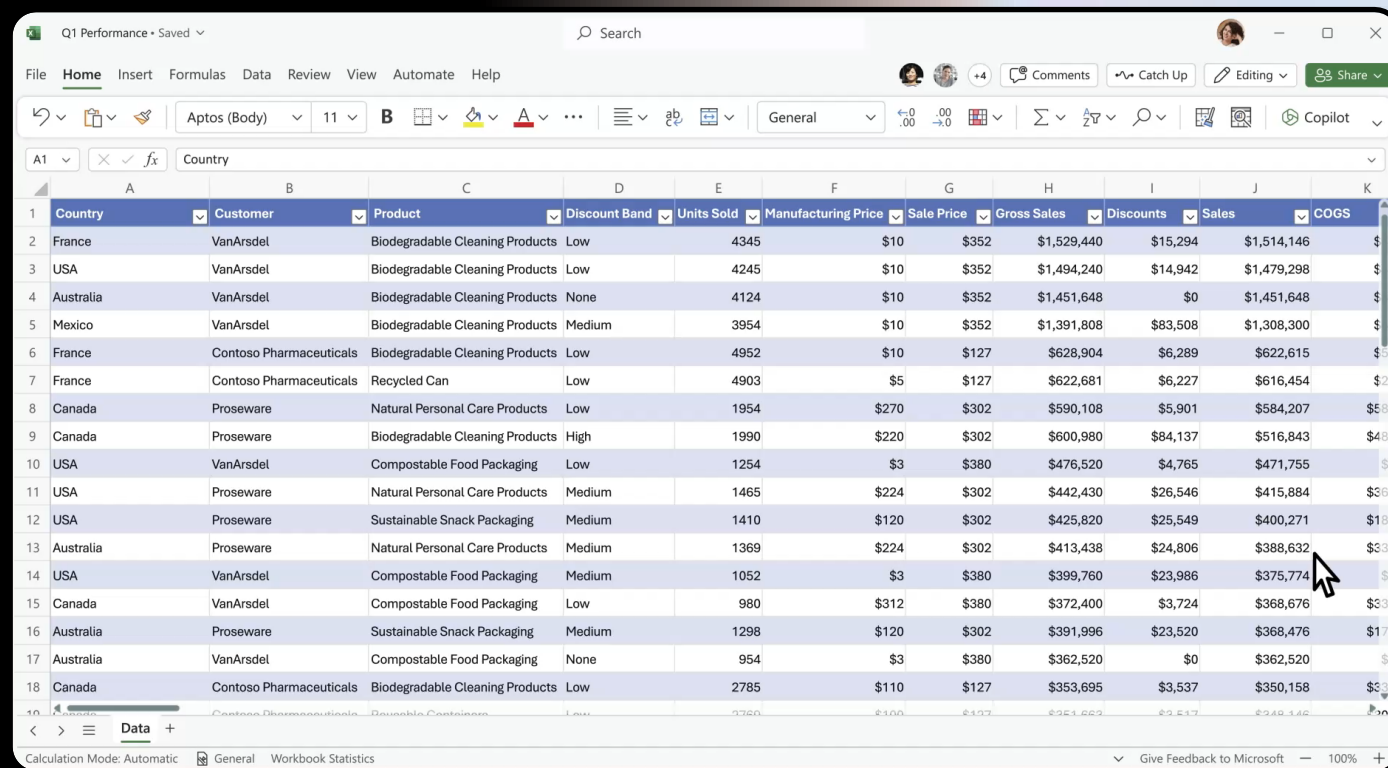
Copilot 与您并肩工作，帮助分析和探索您的数据

您可以询问 Copilot:

按类型和渠道细分销售，插入表格

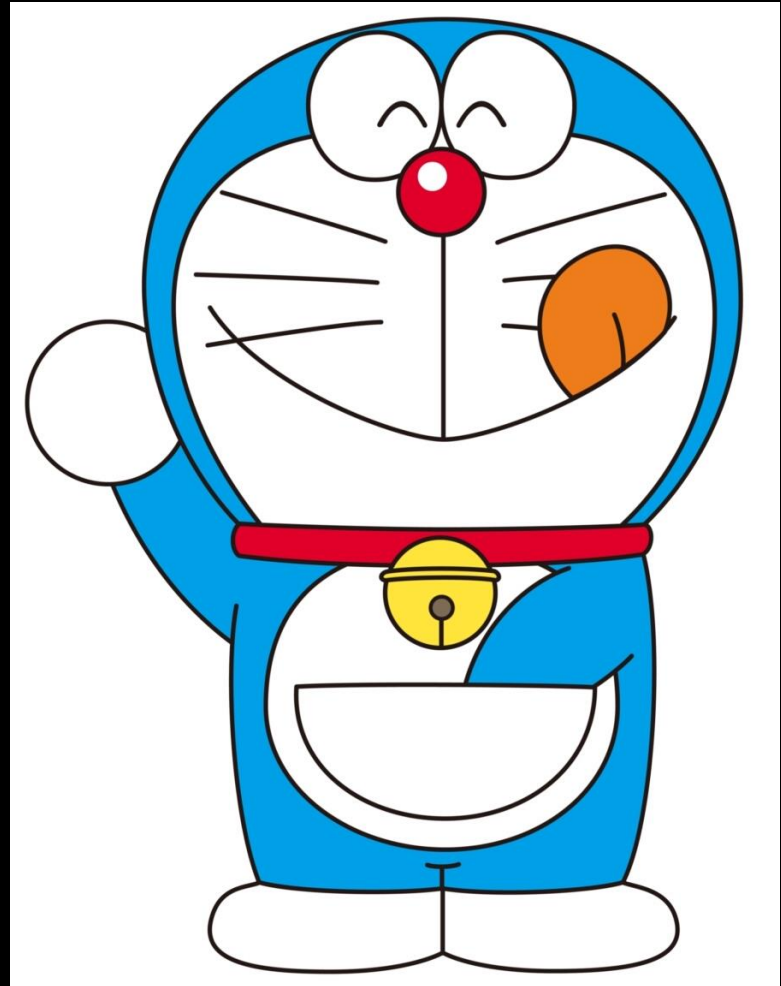
预测 [variable] 的影响并生成图表以帮助可视化

模拟 [variable] 增长率的变化将如何影响我的毛利率



	A	B	C	D	E	F	G	H	I	J	K
1	Country	Customer	Product	Discount Band	Units Sold	Manufacturing Price	Sale Price	Gross Sales	Discounts	Sales	COGS
2	France	VanArsdel	Biodegradable Cleaning Products	Low	4345	\$10	\$352	\$1,529,440	\$15,294	\$1,514,146	\$
3	USA	VanArsdel	Biodegradable Cleaning Products	Low	4245	\$10	\$352	\$1,494,240	\$14,942	\$1,479,298	\$
4	Australia	VanArsdel	Biodegradable Cleaning Products	None	4124	\$10	\$352	\$1,451,648	\$0	\$1,451,648	\$
5	Mexico	VanArsdel	Biodegradable Cleaning Products	Medium	3954	\$10	\$352	\$1,391,808	\$83,508	\$1,308,300	\$
6	France	Contoso Pharmaceuticals	Biodegradable Cleaning Products	Low	4952	\$10	\$127	\$628,904	\$6,289	\$622,615	\$
7	France	Contoso Pharmaceuticals	Recycled Can	Low	4903	\$5	\$127	\$622,681	\$6,227	\$616,454	\$2
8	Canada	Proseware	Natural Personal Care Products	Low	1954	\$270	\$302	\$590,108	\$5,901	\$584,207	\$58
9	Canada	Proseware	Biodegradable Cleaning Products	High	1990	\$220	\$302	\$600,980	\$84,137	\$516,843	\$48
10	USA	VanArsdel	Compostable Food Packaging	Low	1254	\$3	\$380	\$476,520	\$4,765	\$471,755	\$
11	USA	Proseware	Natural Personal Care Products	Medium	1465	\$224	\$302	\$442,430	\$26,546	\$415,884	\$36
12	USA	Proseware	Sustainable Snack Packaging	Medium	1410	\$120	\$302	\$425,820	\$25,549	\$400,271	\$18
13	Australia	Proseware	Natural Personal Care Products	Medium	1369	\$224	\$302	\$413,438	\$24,806	\$388,632	\$33
14	USA	VanArsdel	Compostable Food Packaging	Medium	1052	\$3	\$380	\$399,760	\$23,986	\$375,774	\$
15	Canada	VanArsdel	Compostable Food Packaging	Low	980	\$312	\$380	\$372,400	\$3,724	\$368,676	\$33
16	Australia	Proseware	Sustainable Snack Packaging	Medium	1298	\$120	\$302	\$391,996	\$23,520	\$368,476	\$17
17	Australia	VanArsdel	Compostable Food Packaging	None	954	\$3	\$380	\$362,520	\$0	\$362,520	\$
18	Canada	Contoso Pharmaceuticals	Biodegradable Cleaning Products	Low	2785	\$110	\$127	\$353,695	\$3,537	\$350,158	\$3

Copilot + X



Azure OpenAI Service 客户案例

accenture

avanade

Albusters

ALERT
INNOVATION

AMERICAN FIDELITY
a different opinion

AON

arvato
BERTELSMANN

AUTODESK

BMW GROUP
MINI

CIPIO.ai

CARmax

clevertar

cognizant
intuition engineered

Databook

Downer

ELSEVIER
Health

EY

Farmlands
co-operative

GENIE AI

ILIFAD

Inpris

inworld

Kepro
Making healthcare work.

klaviyo

MATTEL

naïon
nationale nederlanden

pwc

RTL+

SHELL

snelstart

sogeti
Part of Capgemini

soul
machines

STRABAG

TreLent

wipro

WordLift

zammo

世界正在改变

潘多拉的魔盒已经打开

新技术，新起点

行业正在迅速转型

经济发展的驱动力正在演变

世界比以往任何时候都更加紧密相连



予力全球每个人和每个组织
都成就不凡

Empower every person and
every organization on the
planet to achieve more