

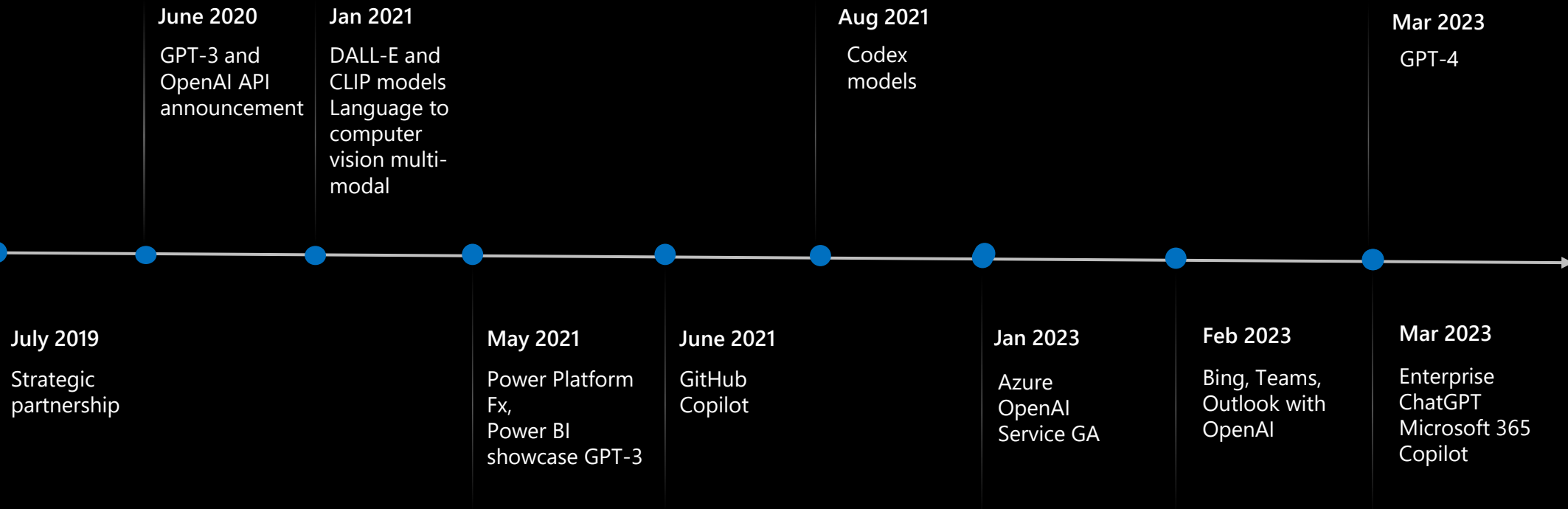


探索人工智能的未来

Azure OpenAI 技术和应用场景进阶

OpenAI 与 Microsoft 战略合作

不仅仅是100亿这么简单

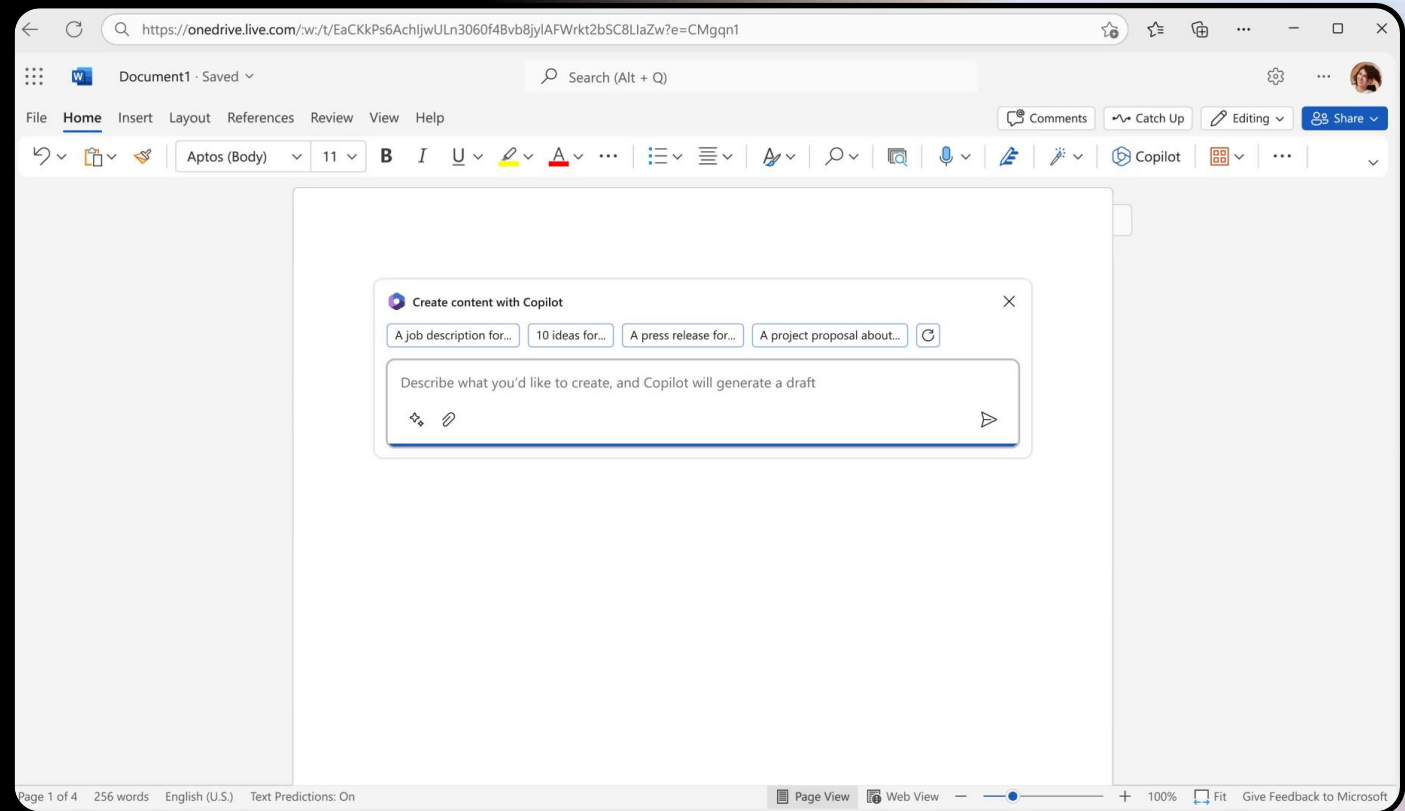


Copilot in Word

Copilot writes, edits, summarizes,
& creates right alongside you.

Some example prompts you can ask **Copilot**:

- Draft a two-page project proposal based on the data from [*document*] and [*spreadsheet*].
- Make the third paragraph more concise. Change the tone of the document to be more casual.
- Create a one-page draft based on this rough outline.

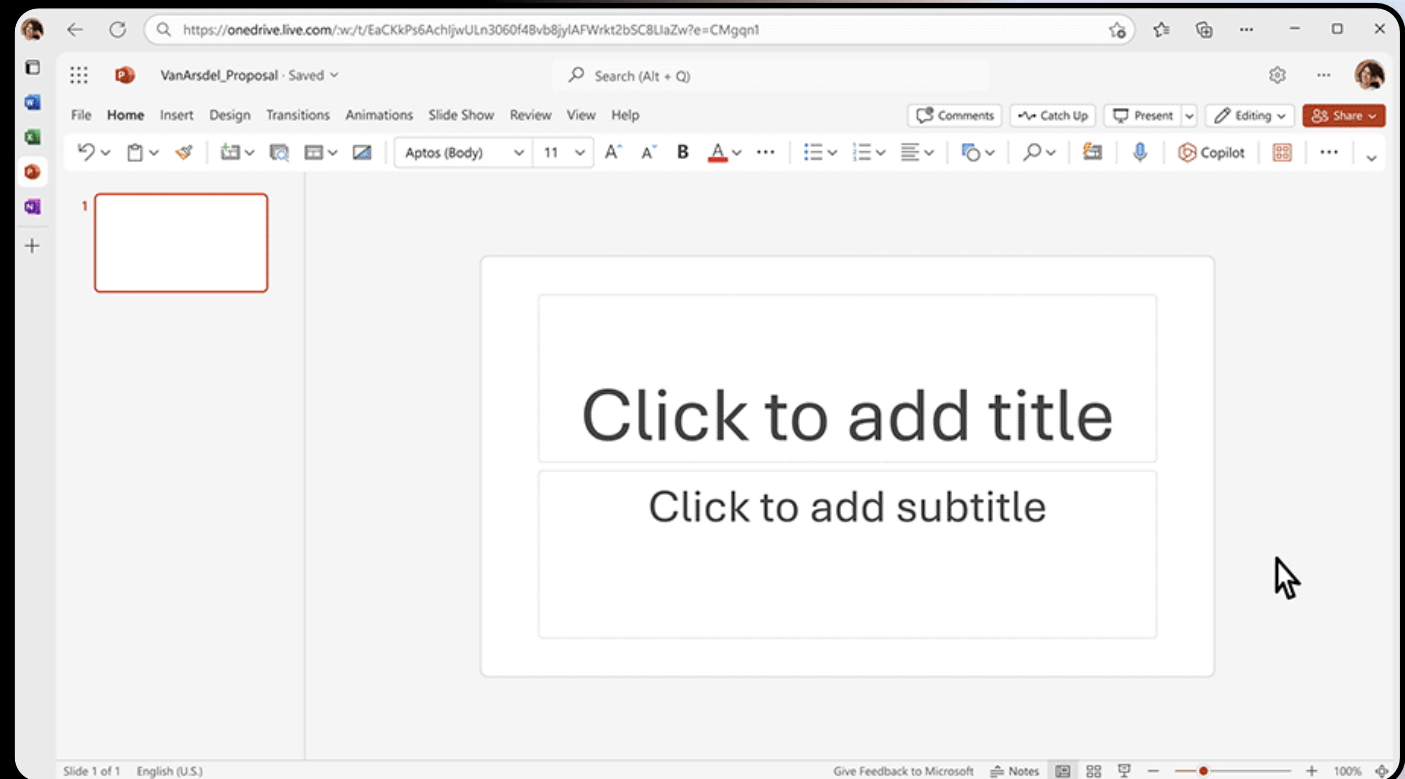


Copilot in PowerPoint

Copilot helps you turn your ideas into stunning presentations.

Some example prompts you can ask **Copilot**:

- Create a five-slide presentation based on a Word document, include relevant stock photos.
- Consolidate this presentation into a three-slide summary.
- Reformat these three bullets into three columns, each with a picture.

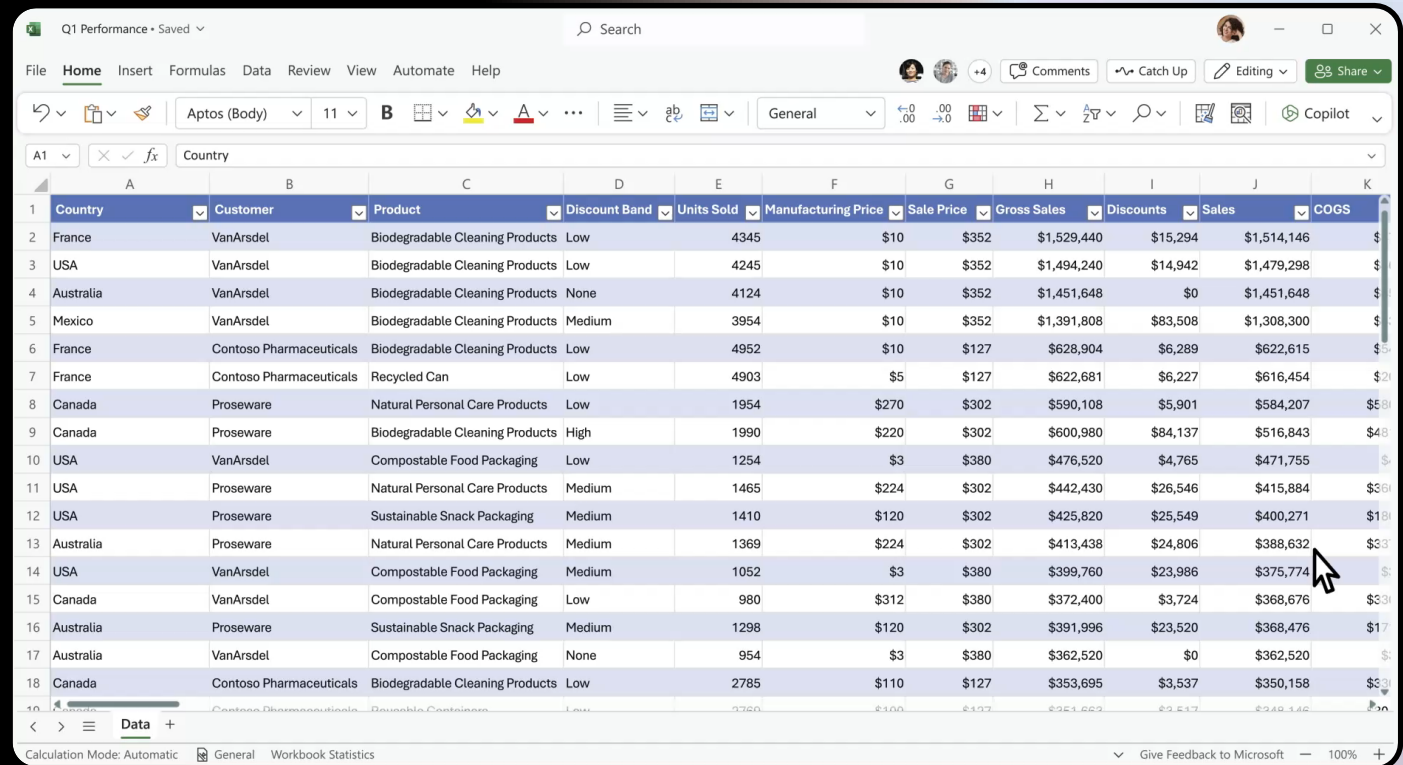


Copilot in Excel

Copilot works alongside you to help analyze and explore your data.

Some example prompts you can ask **Copilot**:

- Give a breakdown of the sales by type and channel. Insert a table.
- Project the impact of [variable] and generate a chart to help visualize.
- Model how a change to the growth rate for [variable] would impact my gross margin.



The screenshot shows the Microsoft Excel interface with a sales data table. The table has columns for Country, Customer, Product, Discount Band, Units Sold, Manufacturing Price, Sale Price, Gross Sales, Discounts, Sales, and COGS. The data is organized by country and product type. The Copilot interface is visible at the top right, showing a search bar and various icons for analysis and visualization.

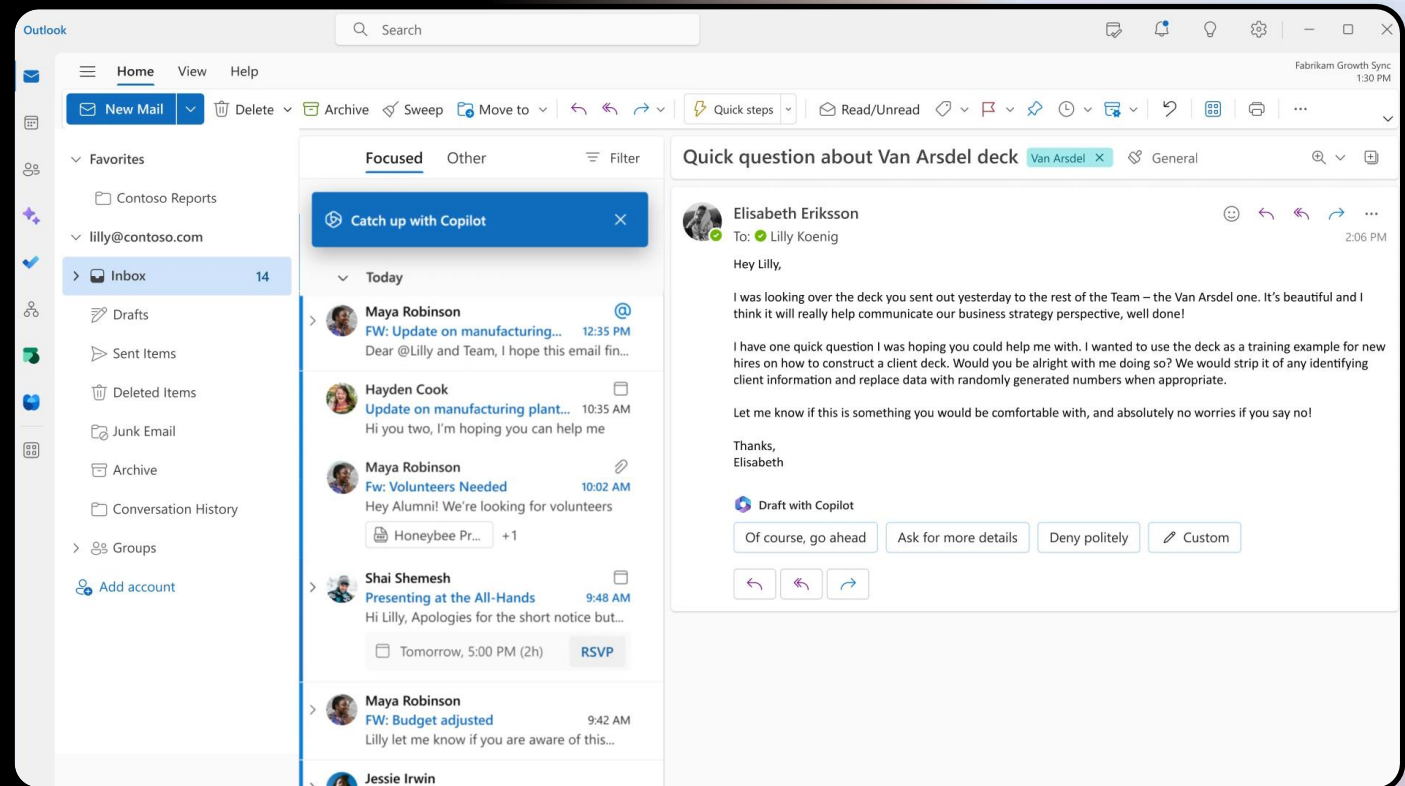
| Country | Customer | Product | Discount Band | Units Sold | Manufacturing Price | Sale Price | Gross Sales | Discounts | Sales | COGS |
|-----------|-------------------------|---------------------------------|---------------|------------|---------------------|------------|-------------|-----------|-------------|-------------|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

Copilot in Outlook

Copilot works with you in cleaning your inbox and messages.

Some example prompts you can ask **Copilot**:

- Summarize the emails I missed while I was out. Flag important items.
- Draft a response thanking them and asking for more details; shorten this draft and make the tone professional.
- Invite everyone to a “lunch and learn” about new product launches next Thursday at noon

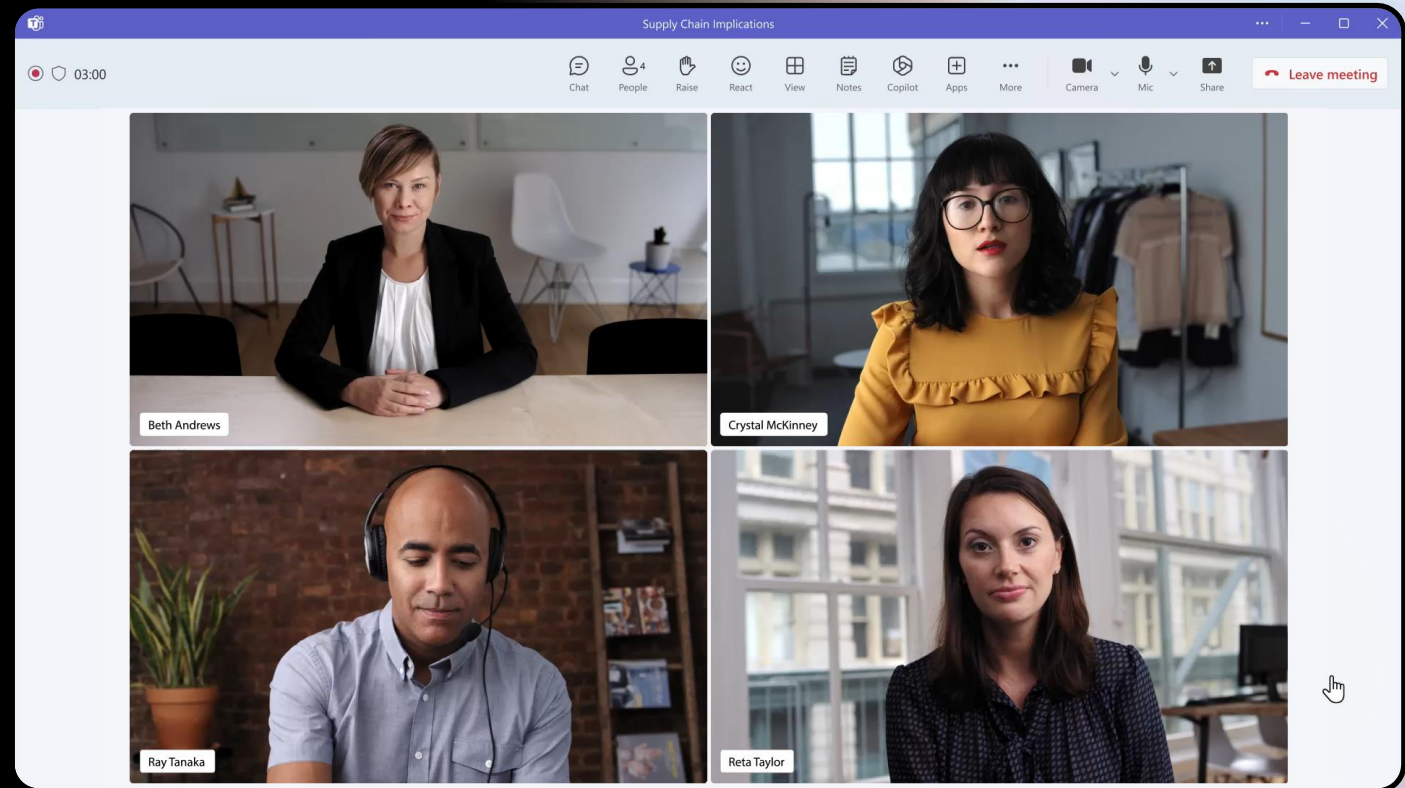


Copilot in Teams

Copilot helps you run effective meetings, get up to speed, organize key discussion points, and summarize actions.

Some example prompts you can ask **Copilot**:

- Summarize what I've missed in the meeting.
- Create a table of pros and cons for [topic].
- What decisions were made, and what are some suggested next steps?



| 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 | toC智能终端情感陪伴型 闲聊机器人 | 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 | 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 |

针对私域数据的ChatGPT

LLM有很强的上下文学习能力，因此Customize的方法也和传统思路不太一样。
Embedding可以满足绝大多数场景的需求

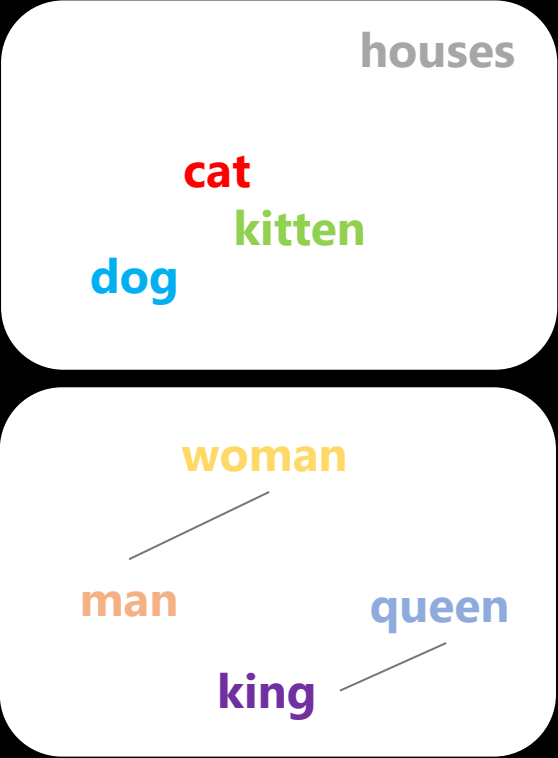
以客服QA Bot为例

1. 使用Embedding模型把知识库(KM)转为矢量，并保存到Vector Database做匹配和搜索
2. 用户对Bot输入一段自然语言
3. 对其做同样的矢量转换 (一样的Embedding模型)
4. 将输入的矢量和Vector Database做近似度计算 (=语义匹配)，抽取出关联度最高的条目/文章。这部分建议搭配Open Source (比如 Redis, FAISS)
5. 返回文章给GPT 作为prompt engineering的一部分，得到自然语言形式的回答
6. 返回结果给用户

Embeddings (search, classify, compare text)

| | 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

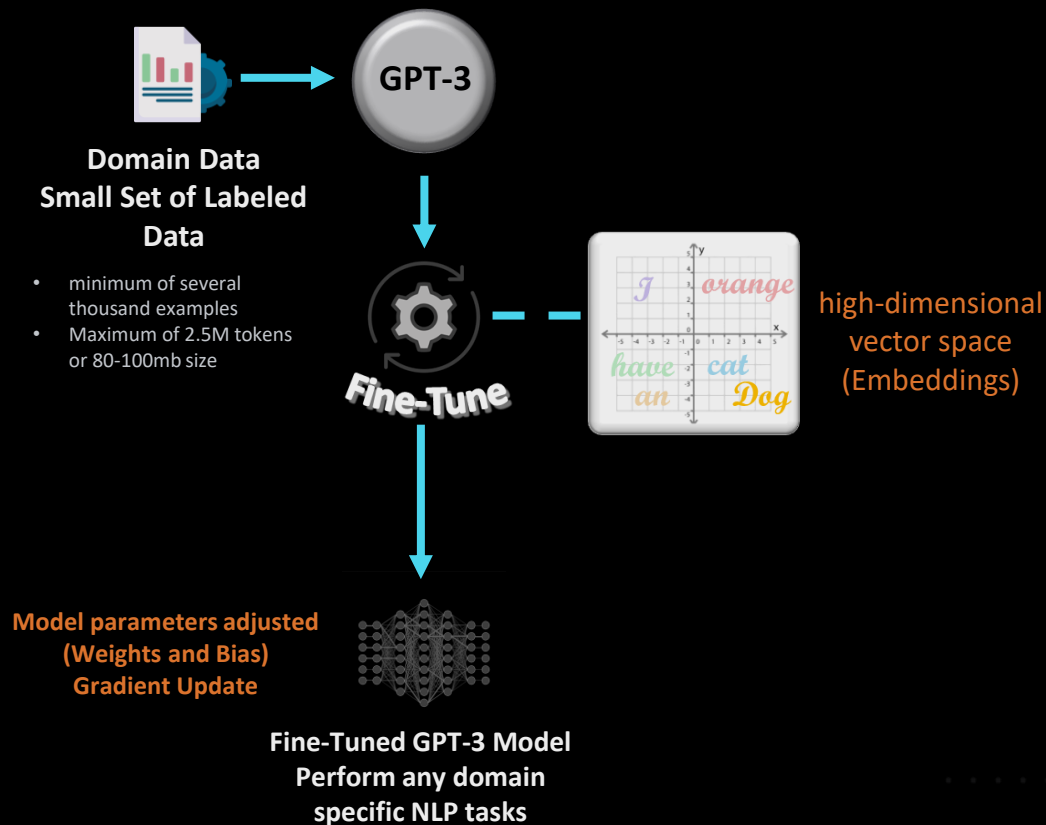
Word embedding

Dimensionality Reduction

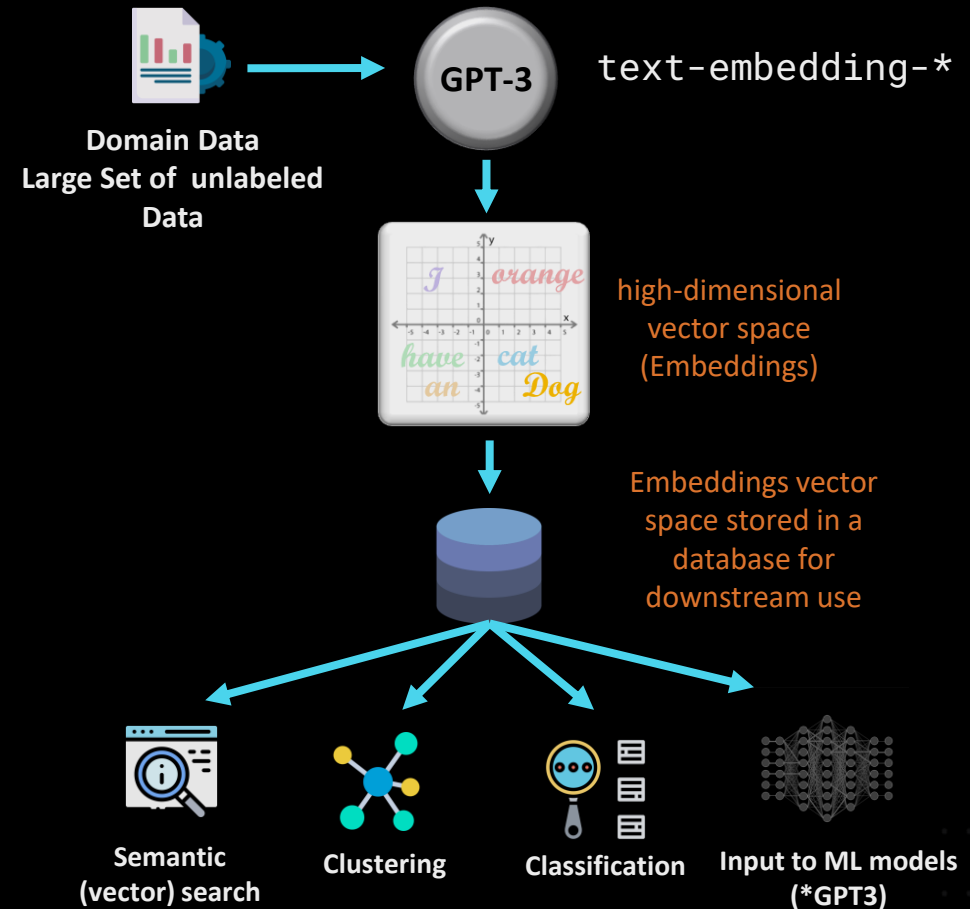
Visualization word in 2D

Model Adaptation with specific domain data

Fine-Tuning



Embeddings



Model End Points

Text Completion Endpoint

text-davinci-003

Base Model Model
Model Family Version

text-curie-001

text-babbage-001

text-ada-001

gpt-35-turbo

Feature-specific end points

text-embedding-davinci-001

Feature

text-embedding-curie-001

text-embedding-babbage-001

text-embedding-ada-002

Code Completion Endpoint

code-davinci-002

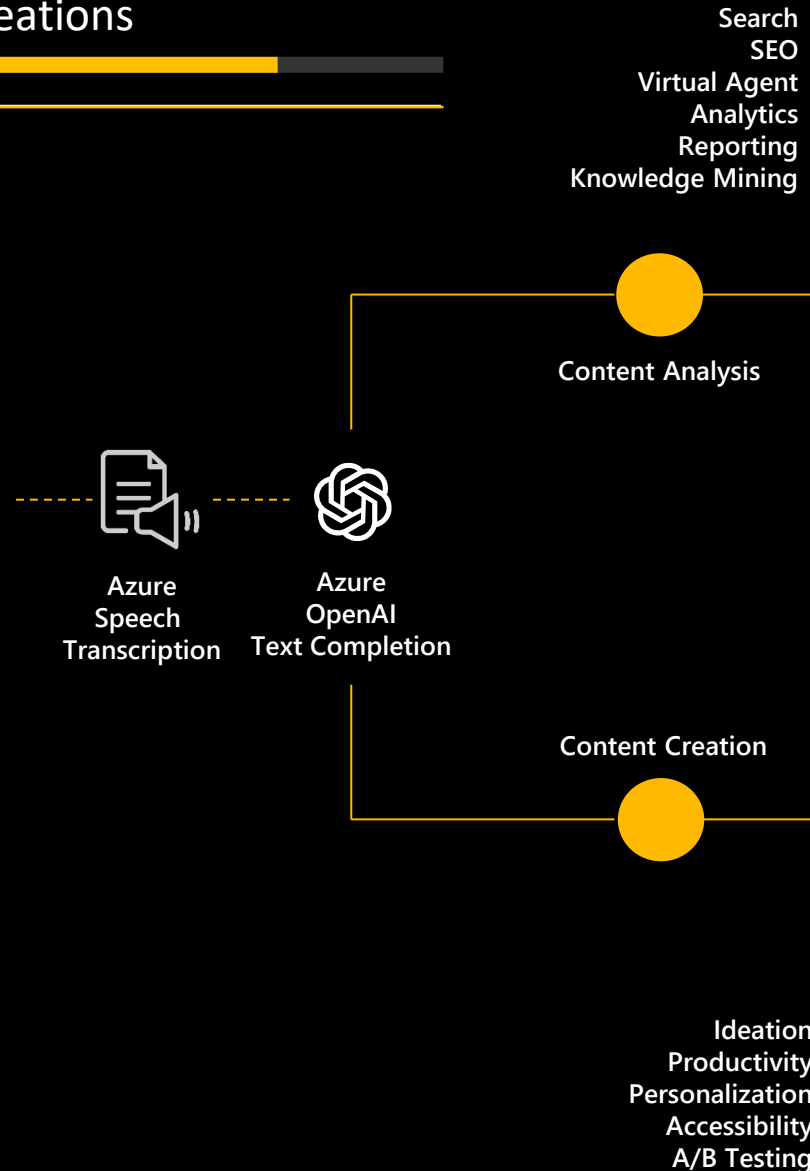
code-cushman-001

Media Example

News Analyses & Article Creations



News Broadcast
(Global Warming)



Topic Classification

Global warming, Deforestation,
Carbon footprint

Entity Extraction

Organizations: IPCC, UNFCCC, Green Peace
Geography: Canada, USA

Key Word Extraction

Human activities, fossil fuels,
earth atmosphere

Question and Answer

What is the Intergovernmental Panel on Climate
Change (IPCC)?

The IPCC is an international organization that studies
climate change and the effectiveness ...

Video summarization

The article discusses about global warming and its effects on
the Earth's atmosphere, wildlife, and human communities. It
states that the primary cause of global warming is

News article generation (or blogs, social media)

Global warming is the gradual increase in the overall
temperature of the Earth's atmosphere, primarily caused ...

Script Generation

Act 1: The show opens with a shot of a beautiful coastal town
Act 2: As the town struggles to cope with the crisis ...

Personalized Content generation (or Advertising)

Simon, as someone passionate about global warming, you are
aware of the urgent threat it poses to our planet ...

Sports & Entertainment Example

Natural language to SQL to surface stats data (Fan Engagement)

"How many points did LeBron James score in 2021?"

```
SELECT SUM(points) FROM
basketball_stats WHERE player =
'LeBron James' AND game_date
BETWEEN '2021-01-01' AND '2021-12-
31';
```



Sports Stats
Database



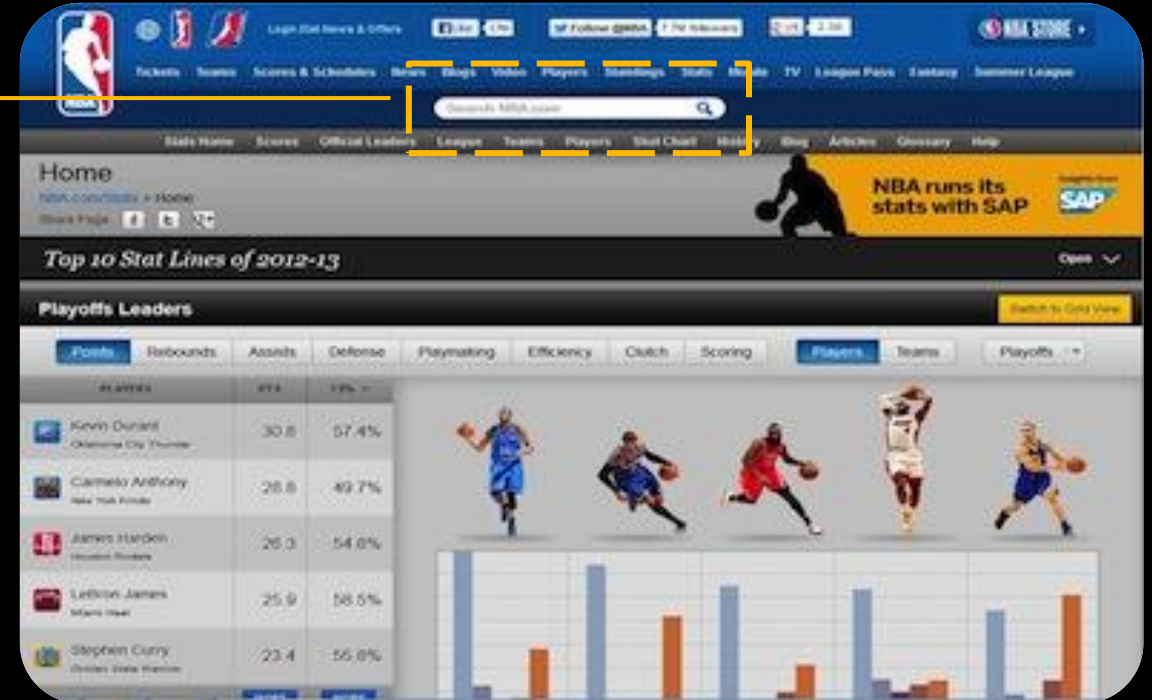
Azure
OpenAI
Codex

basketball_stats

```
25 7 10 1 0 3 82
30 7 4 1 2 3 72
```

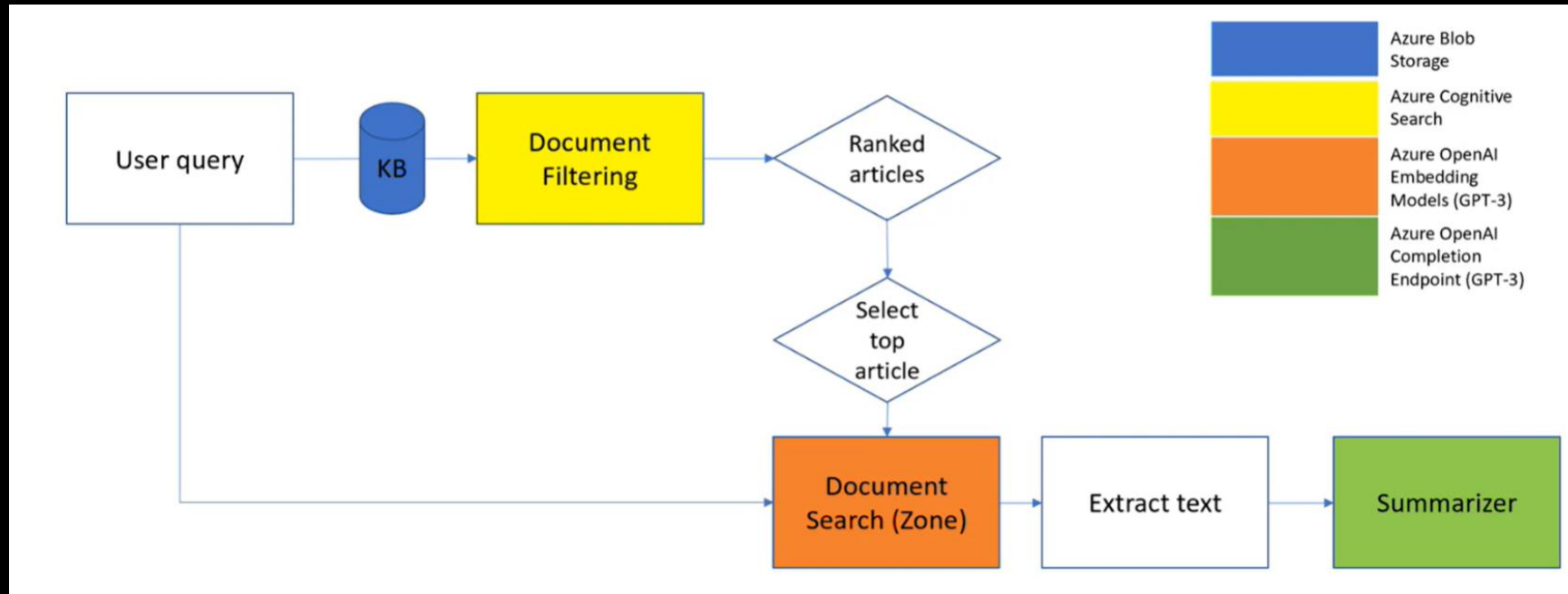
basketball_games

```
110 105 '2021-12-01'
120 115 '2021-12-02'
```



In-house Analyst Example

Query-based Summarization



Use case overview:

- Document search
- Document Zone search
- Text summarization

Advantages of query-based summarization:

1. Shortens reading time
2. Improves the effectiveness of searching for information
3. Removes bias from human summarization techniques
4. Increases bandwidth for humans to focus on more in-depth analysis

Knowledge Base

- CNN/Daily Mail dataset
 - a dataset mainly used for text summarization and question answering tasks
- Schema
 - id: a string containing the heximal formatted SHA1 hash of the URL where the story was retrieved from
 - article: a string containing the body of the news article
 - highlights: a string containing the highlight of the article as written by the article author
- Azure Blob Storage
 - 11.2K articles

Azure Cognitive Search App Example

- On Azure Portal
- Enter a query
- Returns results based on semantic search score
- Can customize what information is retrievable, filterable, etc when designing the index

Azure OpenAI Text Search Embeddings

- Text segmentation
 - Divide document into consumable chunks due to token limitation
 - Paragraph boundary detection (**)
 - Results in "chunks"
- Each chunk gets embedded with doc model ("text-search-curie-doc-001")
- User can send a document specific query – gets embedded with query model ("text-search-curie-query-001")

Using Embeddings for Text Extraction

- Calculate cosine similarity between the query embedding and each chunk's embedding
- Resulting in a ranking of text chunks most relevant to user query
- Extract the most similar chunk

* Refer to Future Work section

Azure OpenAI GPT-3 Completion Endpoint

- Pass extracted text chunk to GPT-3 for summarization
- Outputs summarized text for a specific zone in the document

```
var promptData = new PromptDesign
{
    Prompt = "Summarize the text. Do not make up facts.\n\nText:\n" +
    text.Trim() + "\n\nSummary:\n",
    Max_tokens = 500,
    Temperature = 0
};
```

* frequency_penalty, top_p, etc can all be added easily depending on use case

Cross industry scenarios

- 企业特定领域的知识问答

- Web crawl + QA (Notebook): <https://github.com/openai/openai-cookbook/tree/main/apps/web-crawl-q-and-a>
- Long text + QA: <https://github.com/openai/openai-cookbook/tree/main/apps/file-q-and-a>
- OCR + QA: <https://github.com/ruoccofabrizio/azure-open-ai-embeddings-qna>
- Cognitive Search + QA: <https://github.com/Azure-Samples/azure-search-openai-demo/>

Azure OpenAI与传统AI模型的差异化

更简单，更强大，无限可能



我需要一个可以同时处理多个任务的通用模型
例如:翻译+实体识别+情感分析



我需要生产人工模拟内容，同时保护数据隐私和安全
例如，抽象概括，内容写作，转述，代码



我需要加速原型设计和上市，以满足多种应用场景



我需要一个少量训练或零训练的模型



我想查找之前描述过的解决方案/用户案例，用于头脑风暴或辅助决策