



中信建投证券
CHINA SECURITIES

证券研究报告·行业动态报告

GPT4及ChatGPT相关应用梳理

分析师：阎贵成

yanguicheng@csc.com.cn

SAC编号：S1440518040002

SFC 中央编号：BNS315

分析师：金戈

jinge@csc.com.cn

SAC编号：S1440517110001

SFC 中央编号：BPD352

分析师：于芳博

yufangbo@csc.com.cn

SAC编号：S1440522030001

发布日期：2023年3月17日

核心观点

核心观点：3月15日 OpenAI发布多模态大模型GPT-4，其不仅在语言处理能力上提高，还具备对图像的理解和分析能力。GPT-4商业化进程加快，开放API的同时还发布了在6个不同商业场景的应用落地。随后微软发布了震撼的微软365 Copilot，极大提升office的生产力和交互方式。此前，ChatGPT就已经展示了强大的能力，在文字创造、人机交互、教育、影音、零售等多场景落地应用。我们认为，多模态大模型已在多领域具有专家能力，未来将深度赋能千行百业，改变生产生活方式。

1. **2023年3月15日，OpenAI目前最强大的多模态预训练大模型GPT-4正式发布。**其能够处理文本、图像两种模态的输入信息，单次处理文本量是ChatGPT的8倍，GPT-4的表现大大优于目前最好的语言模型，同时在学术考试中的水平远超GPT3.5。这意味着GPT-4不仅在学术层面上实现了模型优化与突破，同时也展现出了成为部分领域专家的能力。
2. **GPT-4商业化进程加快。**GPT-4在发布时便开放了其纯文本输入的API，这与GPT-3和GPT-3.5的滞后开放不同。同时，GPT-4这次一起推出了6个商业场景的具体应用落地，在优化人机交互体验、提供专业服务、提升组织效能、文化遗产与保护等方面都展现了巨大的潜能，未来有望看到更多场景中的商业化拓展与落地。
3. **ChatGPT已经刮起GPT生态狂潮。**2023年3月1号，OpenAI基于GPT 3.5 Turbo模型开放了ChatGPT API。API收费模式为0.002美元/1000tokens。相较于前一代开放接口GPT 3.5，性能更强的同时，价格下降90%，加速ChatGPT相关应用生态的发展。
4. **应用百花齐放，创造新的生产方式。**微软先后在搜索引擎Bing、企业服务平台Dynamic 365及开发者工具平台Power Platform等接入ChatGPT/GPT-4能力。微软还发布了震撼的Microsoft 365 Copilot，极大提升office的生产力和交互方式。与此同时，在越来越多的企业宣布接入ChatGPT的能力，其中不乏一些已经取得优秀商业化的应用。如Jasper、Quizlet、Shop等，在语言文字创造、人机交互、教育、绘画、影音、零售等多场景落地应用。
5. **我们认为以GPT4/ChatGPT为代表的预训练大模型是这一轮AI浪潮的重要革新，重点推荐关注海内外优秀应用落地情况、AI芯片产业链等投资机遇。**



目录

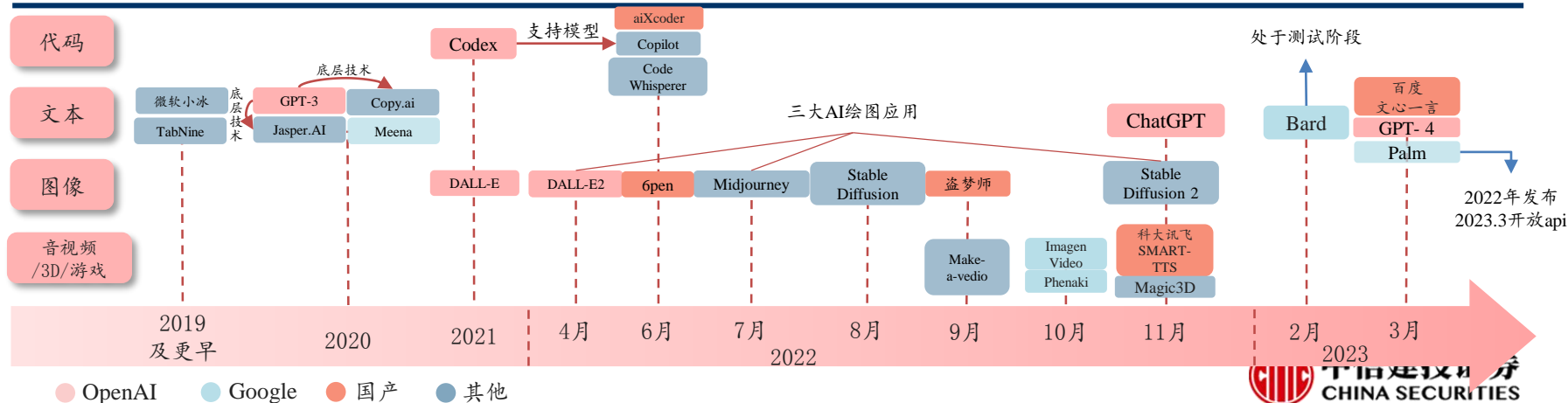
一、ChatGPT：从全民关注的爆款应用走向商业化开放，在多个场景中展现价值

二、GPT4：更强大的多模态大模型，带来更广泛的商业应用可能

1.1 AIGC商业化演进：重点产品时间线

- 随着技术的快速进步，AI生成内容(AIGC)极大地解放了生产力，其广阔的应用场景及市场价值推动着其商业化的快速发展。
- 垂直领域快速落地，爆款产品不断推出。**随着技术不断演进，尤其是Diffusion、Transformer和Clip等模型的提出，AIGC的商业化在多个领域开始落地。其中代码生成、文本生成、图像生成、音视频/游戏等领域已有不少爆款级产品。
- 1) 代码领域，基于OpenAI的Codex模型，Github推出的Copilot具备代码的生成、补全、建议等功能，用户量突破百万级别。
- 2) 文本领域，ChatGPT目前已开放API接口，微软已经在Bing和Dynamic365等应用接入。3月份发布的多模态预训练大模型GPT-4已开放纯文本API，在图片理解、推理和学术考试中展现了强大能力，未来将适用于更广泛的应用场景。
- 3) 图像领域，Stable diffusion等文生图应用已突破千万级用户量，目前是商业模式最成熟的领域，国内也在积极布局。
- 4) 音视频/游戏领域，起步较晚，对模型复杂度和算力要求更高，百度的文心一言大模型已经具备对视频生成的能力。

图：AIGC重点产品时间线



1.2 ChatGPT开放接口， OpenAI商业化进一步推进

- **OpenAI开放ChatGPT API和Whisper API（应用程序接口），允许开发者将ChatGPT集成至他们的应用程序和服务中。**
- 2023年3月1号，OpenAI宣布开放了ChatGPT API，其由自然语言大模型“GPT 3.5 Turbo”支持，开放的API收费模式为0.002美元/1000tokens，较前一代价格下降90%。
- **OpenAI根据API调用的token（分词）输入和输出总数计费。**例如调用输入10个token，输出20个token，则需要支付30个token的费用。同时token总数还需低于模型最大限制，GPT-3.5-turbo-0301版本的限制是4096个token。
- **除此之外，本次OpenAI还开放了Whisper API。**是一种语音转换文字（speech to text）的模型，可以转写或翻译语音，支持包括英语、中文、阿拉伯语、日语、德语、西班牙语等几十种语言，使用Whisper API的费用为每分钟0.006美元。
- **目前OpenAI已经开放了包括ChatGPT API和Whisper API在内的多个API接口，逐步推进商业化进程：**

OpenAI开放的语言类API包括用于对话的GPT-3.5-turbo（多段对话）、InstructGPT（一问一答），其中InstructGPT根据其性能可以分为Ada、Babbage、Curie、Davinci（由弱至强），这四个模型还提供针对具体场景的微调接口，其中的Ada还可以作为研究的基础Embedding模型被调用。开放的多模态模型为Dall E，根据处理图像的像素不同进行收费。开放的音频模型为Whisper。

表：OpenAI 开放的相关模型

应用	具体功能	模型名	定价 (美金/1000tokens)	应用	具体功能	模型名	定价 (美金/1000tokens)	应用	具体功能	模型名	定价										
语言模型	对话	GPT-3.5-turbo		语言模型	Fine-Tuning	Ada	0.0004 (训练) 0.0016 (调用)	多模态模型	文生图	Dall·E	0.020美金/ image (1024×1024)										
		0.002				Babbage	0.0006 (训练) 0.0024 (调用)				0.018美金/ image (512×512)										
		Ada	0.0004			Curie	0.0030 (训练) 0.0120 (调用)				0.016美金/ image (256×256)										
		Instruct GPT	Babbage			0.0005	Davinci				0.0300 (训练) 0.12 (调用)	音频模型	语音识别	Whisper	0.006美金 / minute						
		Curie	0.0020			Embedding	Ada				0.0004										
		Davinci	0.0200																		
资料来源: OpenAI, 中信建投																					

资料来源：OpenAI，中信建投

1.3 微软与OpenAI深度合作，依托前沿AI技术提升产品能力

- 微软作为OpenAI最为重要的合作伙伴和投资人，其旗下产品与OpenAI的相关模型深度融合。

- 1.微软旗下的搜索引擎结合了GPT4的技术，将搜索、创造、聊天结为一体，多方位地满足用户个性化的需求；
- 2.微软旗下的企业服务平台Dynamic 365是融合CRM、ERP等功能的SaaS服务，其结合ChatGPT的技术推出Dynamics 365 Copilot，为销售、服务、营销到供应链等跨业务的人员提供帮助；
- 3.微软旗下的开发者工具Power Platform，结合ChatGPT技术推出，Power Virtual Agents和AI Builder，分别提供构建聊天机器人的低代码方法和面向开发人员的生成AI功能。

- 微软产品在融合了生成式AI技术后展现出更强的产品能力和市场竞争力

1. 以Dynamic 365 Copilot为例，在原先的工作场景中，营销人员需要每天花费66%的时间进行邮件回复（根据Futurum研究），而Dynamics 365 Copilot的推出就有望大大减轻这方面的工作量。
2. 产品能力的提升也带来了广泛的市场关注度和认可度，以微软Bing为例，Bing接入ChatGPT一个月后，微软公布进展称，经过多年的稳步发展，加上数百万新Bing预览用户的推动，Bing每日活跃用户已突破1亿。在新Bing预览版的数百万活跃用户中，大约三分之一是Bing的新用户。

图：Bing基于用户指令生成产品对比信息表格

图：微软Power Platform平台与AI结合相关产品时间线

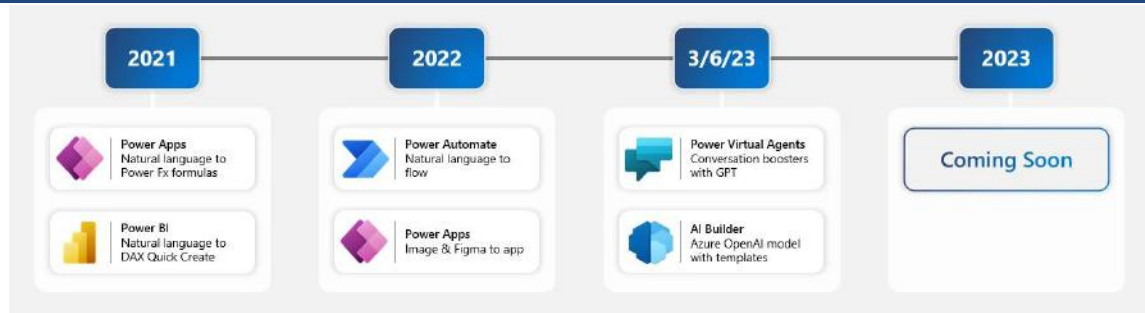


Create a table comparing the specifications of the M2 MacBook Air and Surface Laptop 4.

✓ Searching for: M2 MacBook Air specifications
✓ Searching for: Surface Laptop 4 specifications
✓ Generating answers for you...

Sure, I can create a table comparing the specifications of the M2 MacBook Air and Surface Laptop 4. Here is the table:

Specification	M2 MacBook Air	Surface Laptop 4
Starting Price	\$1,199	\$999
Dimensions (inches)	11.97 x 8.46 x 0.44	13.0" 12.1 x 8.8 x 0.57, 15" 13.4 x 9.6 x 0.58
Weight (pounds)	2.7	13.0" 2.78 (Intel), 2.84 (AMD), 15" 3.4 (Intel), 3.48 (AMD)
Display	13.6-inch Liquid Retina, Wide Color (P3), True Tone, 2560 x 1600 resolution (AMD)	13.0" or 15" PixelSense, 2256 x 1504 (13.0") or 2496 x 1664 (15.0") resolution, 500:1 (13.0") or 500:1 (15.0")

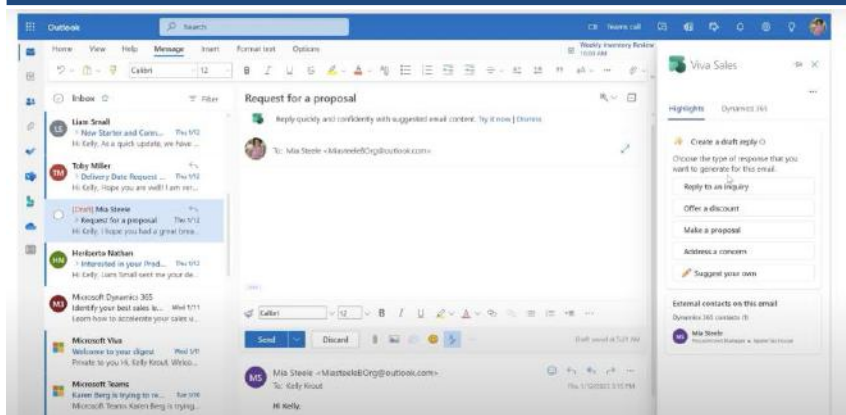


资料来源：OpenAI，中信建投

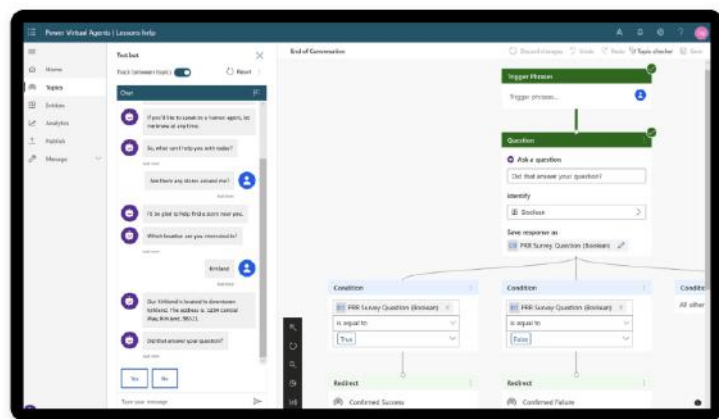
1.4.1 B端应用：Dynamic365&Power Platform

- **Dynamic365**是微软推出的企业服务平台，其中于2023年3月9日推出的Dynamics 365 Copilot结合了ChatGPT相关的AI技术，用于CRM和ERP。目前处于预览阶段，其可以为销售、服务、营销到供应链等跨业务的人员提供帮助，例如实现生成回复客户信息的电子邮件，为团队创建会议摘要，为营销活动提供灵感等。
- **Power Platform**是微软旗下的开发者工具，其中于2023年3月6日推出的Power Virtual Agents和AI Builder都已经使用了OpenAI的GPT语言模型技术。其中Power Virtual Agents通过聊天机器人的形式实现低代码编程，比如将AI聊天机器人指向企业的特定资源，如网站或者知识库，简单配置便可使用，无须额外编写代码。而AI Builder结合了ChatGPT API（称为Azure OpenAI服务）的生成AI功能，允许开发人员使用GPT模型生成文本，包括汇总报告、生成内容创意等等。

图：使用Dynamic 365 Copilot辅助回复邮件



图：利用Power Virtual Agents便捷构建聊天机器人



资料来源：微软，中信建投

1.4.2 B端应用：Jasper

- **Jasper：结合GPT-3，打造多款产品**

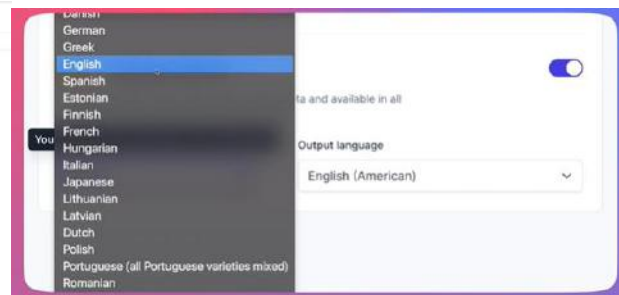
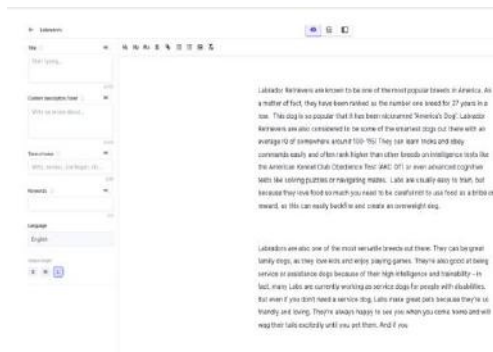
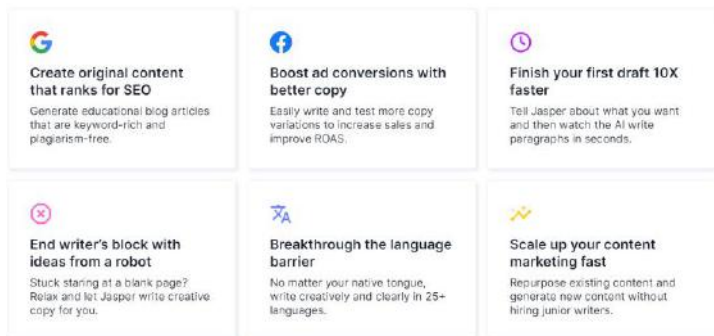
截至2022年12月底，Jasper付费用户达到7万，2022年年收入7500万美金，估值已达15亿美金。旗下有多款产品接入GPT-3进行商业性文字创作：

1. Jasper Starter能够利用预定义的模板完成摘要、产品描述、博客文章。
2. Boss Mode是一款写作助手，能够实现灵活便捷的商业文案写作。通过使用Boss Mode，用户仅需要自己编辑20%的内容，Jasper会完成80%的内容。
3. Jasper for Business是专门为B端打造的服务，能够定制化生成文案。用户能够创建新的自定义工作流程，开发专门满足其需求的写作助手。
4. Jasper Chat是一款起到激发想法和修改内容功能的聊天机器人，而Jasper Art能够为广告、博客文章生成图片。

图：Jasper功能简述

图：Jasper将关键词和提示转换为文章

图：Jasper能够翻译25种语言



1.4.3 B端应用：Brex&Slack&ChatSpot.ai

- **Brex：与OpenAi合作，打造便捷化的金融信息平台**

Brex是一家美国的金融科技平台。2023年3月7日，Brex宣布推出基于OpenAI技术的先进人工智能工具。通过Brex Empower平台，客户可以在聊天界面中，使用自然语言来描述要求，获得相应的财务相关信息或是更为复杂的财务分析。

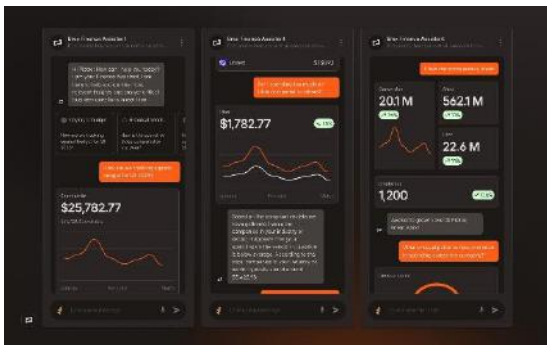
- **Slack：基于ChatGPT能力，助力日常办公的效率提升**

Slack是Salesforce旗下的团队协作工具，与Teams类似。2023年3月7日，Slack宣布引入OpenAI的ChatGPT，用户可以在应用程序中直接对ChatGPT提问来获得问题的答案。ChatGPT将协助用户进行内容理解、总结与归纳，进而实现工作效率的提高。这一功能最初为OpenAI内部团队的协作工具，现在已经向所有的Slack用户开放。

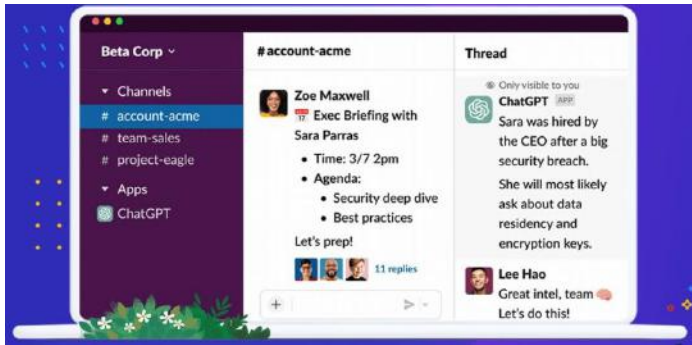
- **ChatSpot.ai：将ChatGPT等AIGC技术与CRM结合，辅助实现多类任务**

HubSpot是一个著名的营销、销售、服务平台（CRM），其推出了一种新的生成AI工具ChatSpot，将HubSpot自己的CRM技术与OpenAI的ChatGPT、DALL-E 2和谷歌文档应用程序结合，可以用自然语言完成多类任务，如调取数据，生成报告等功能。

图：Brex产品示意图



图：Slack中的ChatGPT相关应用



1.4.4 B端应用：HireStack&Pencil

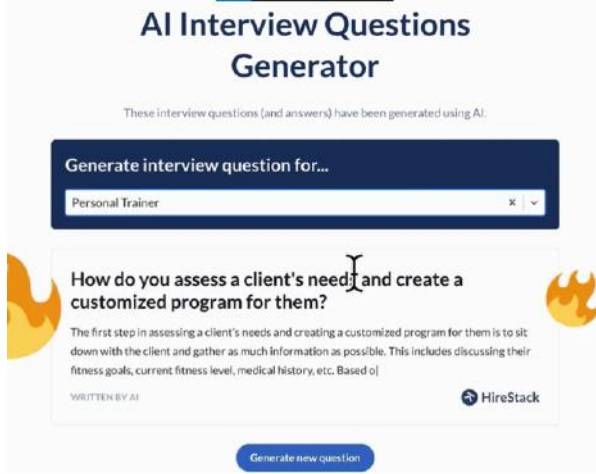
• HireStack：结合GPT-3提升面试效率

HireStack是一家人力资源公司，它利用GPT-3生成面试问题以及对应的答案。参考AI给出的回答，HireStack能够在面试时评估应聘者的专业能力以及对应职位的匹配度，相比于传统的人力资源公司，AI能够极大提升HireStack工作效率。

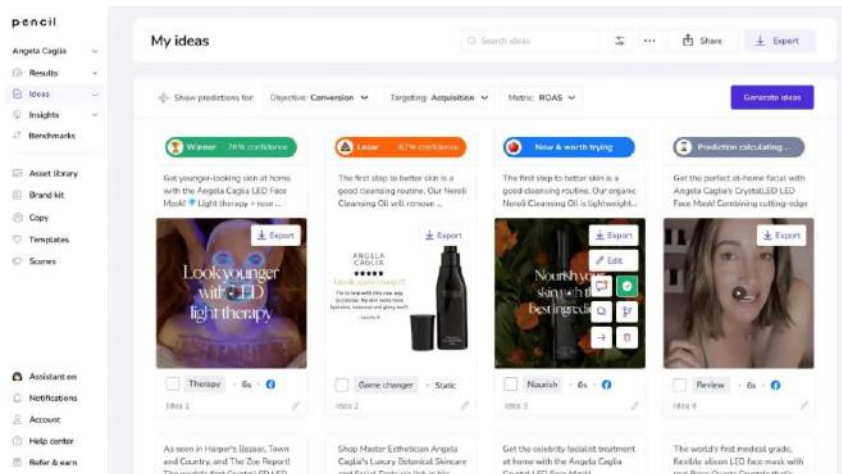
• Pencil：基于GPT-3生成多样化广告

Pencil可以生成效果良好、富有创意且符合品牌要求的广告文案和视频。GPT-3 使得 Pencil 不仅能够一般性地描述产品，而是能够围绕产品进行创意营销。例如，为打造一个手电筒广告，Pencil在广告中首先展现出在黑暗中工作是十分危险的，进而突出手电筒的重要性。更具创意的解决方案将使得Pencil的广告对潜在客户更具吸引力。

图：HireStack利用GPT-3生成面试问题和答案



图：Pencil基于一款产品生成多种类型的广告



1.4.5 C端应用：Snapchat&Discord

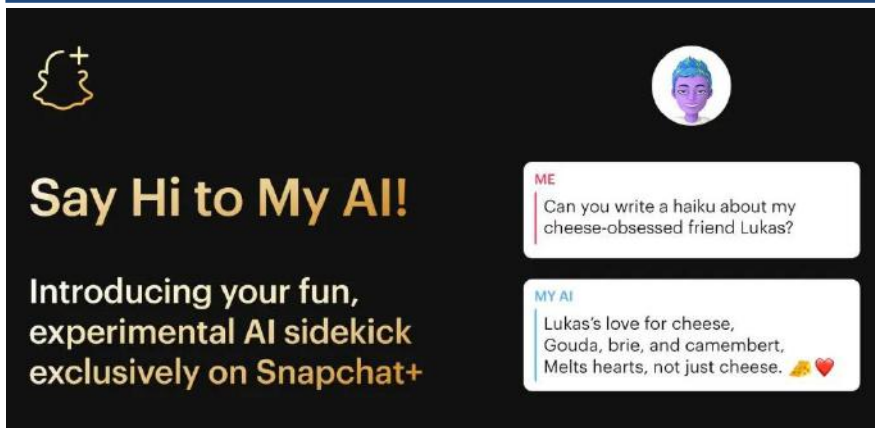
- **Snapchat：基于ChatGPT API推出聊天机器人My AI**

Snapchat是一款由斯坦福大学学生开发的图片分享软件。2023年2月28日Snapchat宣布推出近日基于OpenAI的ChatGPT API的人工智能聊天机器人“My AI”。目前，这一功能仅对订阅了Snapchat+的高级用户开放，订阅费用为每月3.99美元。目前，付费用户数量为两百万，只占公司用户总数的一小部分。公司表示，在不久的将来，平台的所有客户都可以免费或付费使用My AI。

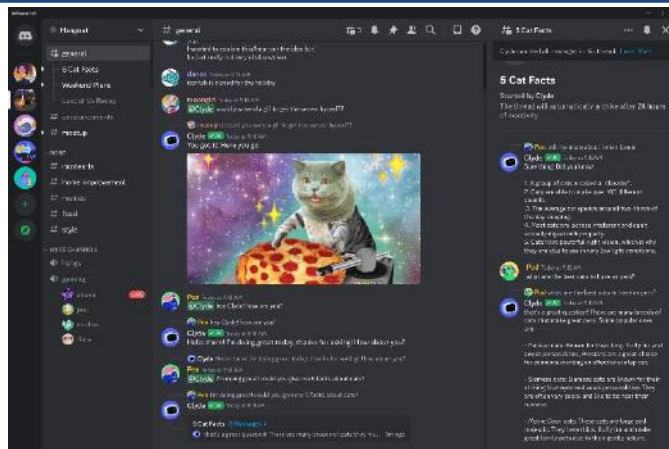
- **Discord：基于ChatGPT API推出聊天机器人Clyde**

Discord是一个游戏聊天应用与社区，每月有3000万用户在社区中体验AI相关的应用。2023年3月10日，Discord宣布引入OpenAI API推出了聊天机器人Clyde，其可以加入到Discord的不同频道中（通过输入“@Clyde”就可以在任意频道中与Clyde进行对话），实现对话总结、基于对话中的问题作答、歌单推荐、与用户进行互动游戏等方面的功能。

图：Snapchat MY AI



图：Discord Clyde



1.4.6 C端应用：Bing

- **Bing：整合Open AI新一代技术，带来搜索引擎范式转移**

2月8日，由ChatGPT和GPT-3.5提供支持的全新搜索引擎Bing正式亮相。新版Bing在传统的搜索外还有对话窗口，可向它提问或是征求它的意见。例如用户可以向其询问参加音乐节的建议。

除此之外，新版Bing还可以进行创作，例如根据提示帮助用户生成从电子邮件到社交媒体的文本，或是表格化的输出帮助用户更好对比产品的差异。

此外，Bing在不断迭代，3月7日，Bing发布了具有三个风格的新版产品以更好满足用户个性化的需求。此外，移动版的Bing也在11日发布，用户在手机上即可使用最新版本的Bing。我们认为，新Bing不同于传统的搜索引擎，将“搜索+创作+聊天”融为一体，能够更加灵活地满足用户多样化需求，或将带来搜索引擎的范式转移。

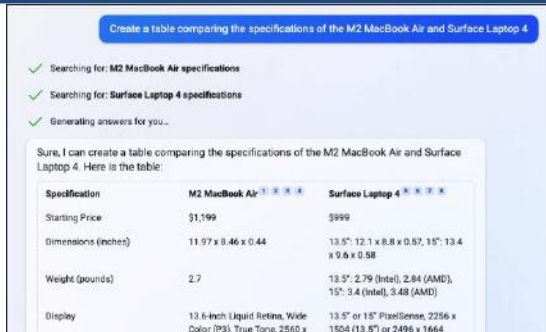
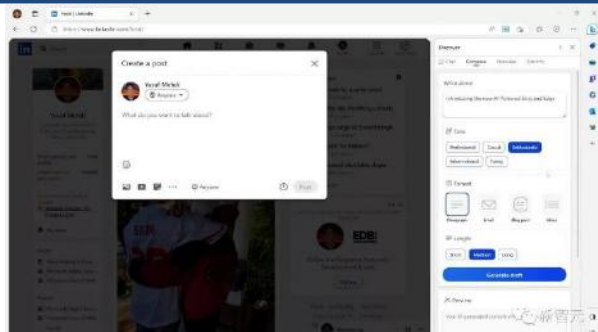
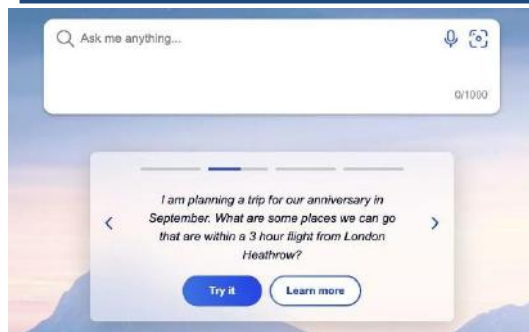
- **新Bing下载量激增，日活已突破1亿**

Bing接入ChatGPT一个月后，微软公布其进展称，经过多年的稳步发展，加上数百万新 Bing 预览用户的推动，Bing 每日活跃用户已突破1亿。在新 Bing 预览版的数百万活跃用户中，大约三分之一是 Bing 的新用户。

图：新Bing界面包含对话框

图：利用Bing生成社交媒体的文本

图：Bing生成表格对比产品信息



1.4.7 C端应用：DuckAssist&Poe

• DuckDuckGo：与OpenAI合作，推出DuckAssist

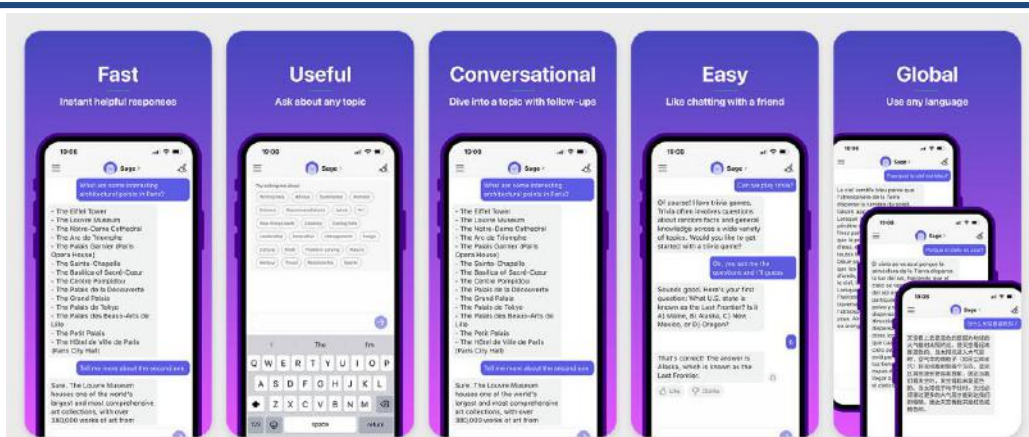
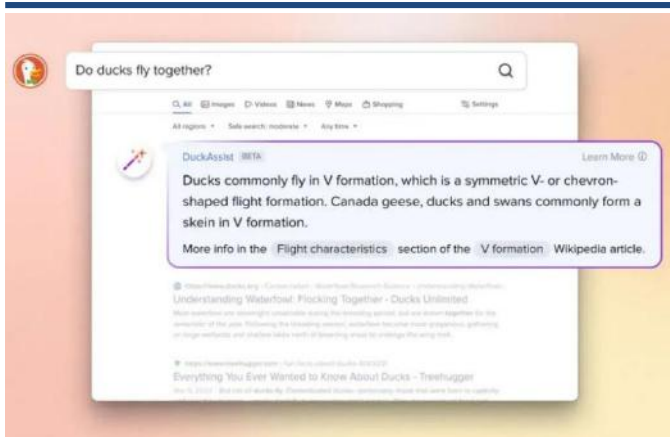
DuckDuckGo是一个主打隐私保护的互联网搜索引擎，于2023年3月9日推出了测试版 DuckAssist，通过 AI 工具来帮助用户快速找到答案。DuckAssist 的设计目标是从维基百科等类似来源中，总结出用户所需要的答案，并以自然语言的方式传达给用户。DuckDuckGo 邀请所有用户对DuckAssist的功能进行测试，且不需要注册用户即可直接使用相关功能，如果测试进展顺利，它将在“未来几周内”正式添加到浏览器中。

• Quora：基于不同的语言模型，推出了聊天机器人Poe

Quora是一个专业问答平台/知识问答社区，类似于美国版的知乎。Quora推出了新的AI聊天机器人应用Poe，最早在2022年12月亮相，但在2023年2月全面发布。所有iOS用户都可以通过电话号码和电子邮件注册账户，然后就能在三种不同的AI聊天机器人中切换，获得相应的聊天结果。具体的聊天机器人包括Sage、Dragonfly（均由OpenAI支持）、Claude（由支持Anthropic）。目前Poe作为独立产品推出，但若Poe上的相关内容达到一定的质量标准，就会被发布至Quora网站上，供每月4亿访问者快速查询。

图：DuckAssist界面示意图

图：Quora POE



1.4.8 C端应用：Quizlet&Instacart&Shop

• Quizlet: 结合ChatGPT API, 推出AI导师Q-Chat

Quizlet是一个全球性的学习平台, 有超过6000万的用户。结合ChatGPT API, Quizlet于2023年2月28日推出了AI导师Q-Chat, 其可以基于相关学习材料提出自适应问题, 生成有趣的聊天内容来与学生进行互动。

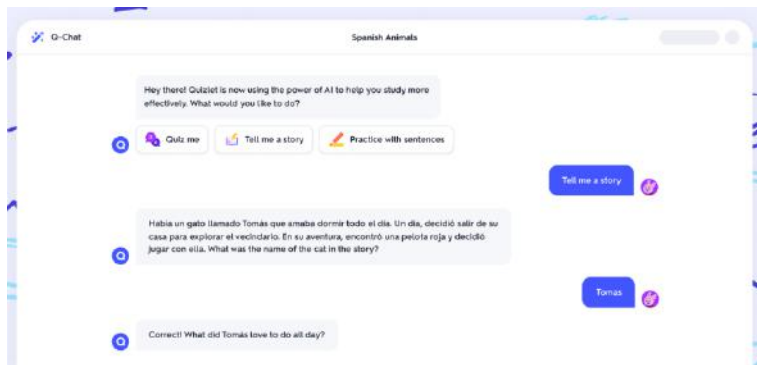
• Instacart: 基于ChatGPT和相关AI技术, 推出Ask Instacart功能

Instacart是一家美国的生鲜杂货电商平台。Instacart利用ChatGPT和自研的AI技术, 结合平台上超过75000个零售合作伙伴店铺的产品信息, 帮助客户实现他们的开放式购物目标, 例如可以回答客户的“我的孩子们午餐吃什么健康的食物?”一类的开放性问题, 这一功能名为Ask Instacart, 将在今年晚些时候推出。

• Shopify: 结合ChatGPT, 提供消费个性化推荐服务

Shop是Shopify的消费者应用程序, 有上亿消费者在程序上寻找品牌和产品信息。通过与ChatGPT的结合, Shop购物助手会根据用户文字描述的需求, 扫描数百万商品快速找到买家所需, 提供个性化推荐, 实现物流流程的简便化、高效化。

图: Quizlet Q-Chat生成趣味问答与学生交互



提问什么是适合孩子的健康午餐

给出食物推荐和购买链接

图: Ask Instacart

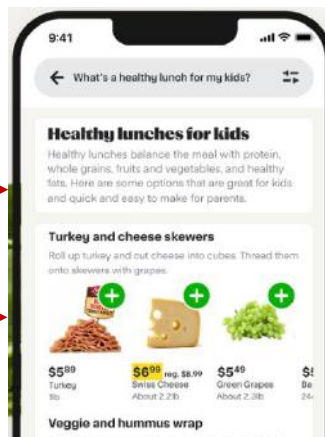
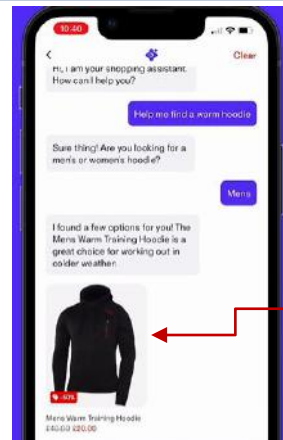


图: Shopify中的shop协助找到买家所需



提出需要寻找的服装需求

根据用户需求推荐相关产品

1.4.9 C端应用：Anania&Tabnine

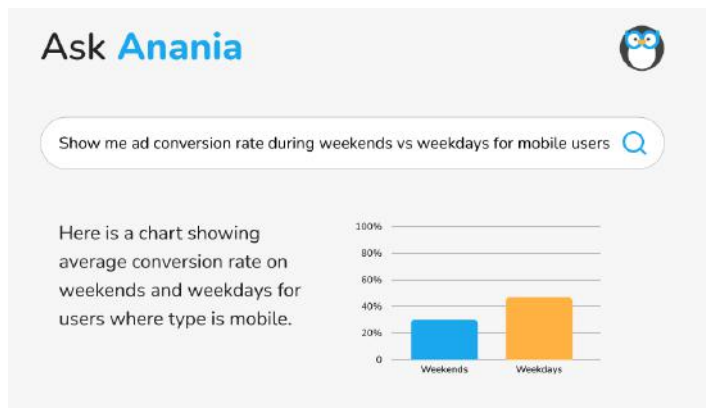
- **Anania：结合多种语言模型进行数据分析**

Anania 是一款数据分析软件，GPT3是Anania使用的AI模型之一。其可以通过简单、自然的英语对话来分析用户的数据。用户可以通过它对数据进行搜集、分析、可视化和导出。例如用户输入“向我展示出周内和周末用手机进行对话的用户比例的差异”，Anania即可快速生成对比图表。

- **Tabnine：结合GPT-2生成代码**

Tabnine使用生成AI技术来预测并根据上下文和语法来推测用户的下一行代码。在编写代码时，Tabnine可以为用户推荐其预测的整行代码。或用户输入注释，Tabnine 则会为其推荐正确的代码。

图：输入句子Anania即可作出图表



图：输入注释Tabnine即可自动生成代码

The screenshot shows the Tabnine code completion interface. It displays a Python code snippet for fitting a random forest model. The code is as follows:

```
# fit random forest model
rf = RandomForestRegressor()
rf.fit(X_train, y_train)

# predict labels of test set
y_pred = rf.predict(X_test)

# calculate mean squared error
```

Annotations on the right side of the code indicate that Tabnine automatically generated the relevant code for the provided comments.

1.4.10 C端应用：Jarvis Lyrics&Rationale

• Jarvis Lyrics：基于GPT3的歌词生成工具

Jarvis是歌曲创作的伴侣。它能够根据任何艺术家、流派、标题和歌词的提示生成歌词建议。然后，用户可以根据建议创作自己的歌词。该工具利用随机输入的语句或创意故事生成的歌词使得艺术家获得更多灵感。

• Rationale：基于GPT3的决策助手

用户只需输入无法决定的选项，Rationale就会根据 GPT 和上下文学习算法列出该决定的优缺点、生成 SWOT 分析或进行多标准分析以帮助用户权衡选择。通过考虑所有相关因素和用户的背景，Rationale可以帮助用户做出理性的决定。

图：输入歌手名称即可使用Jarvis生成相关歌词

图：用户输入“吃苹果”后Rationale列出该决定的优缺点

The Jarvis Lyrics interface features a form on the left for input and a list of generated lyrics on the right. The input fields include 'Fill in one or several fields to let Jarvis generate new lyrics.' with sub-fields for 'Drake' (Artist), 'Genre (optional)' (dropdown), and 'Title (optional)'. Below these is a 'Start lyrics with' section with a text area for 'Type some lines and Jarvis will continue it (optional)'. At the bottom left is a 'Chaos' slider and a 'Reset' button. A large blue 'Generate lyrics' button is at the bottom. The right side displays five lines of generated lyrics: 'Blood pressure dropping, choking on the evil', 'Leaving demons on the couch criticising', 'Always catching feels like I'm allergic', 'Too deep in a conversation and I don't want it to end', and 'Mystic with the pen, writing my name in'. The last line is 'I'm trying to get out but they keep on shutting me down'.

The Rationale interface shows a decision analysis for the input 'I will eat an apple a day.' At the top, a dashed box contains the input text with a red arrow pointing to it from the caption '用户无法确定的事项'. Below the input, there are two columns of analysis: 'Pros 优点' (Pros) and 'Cons 缺点' (Cons). The 'Pros' column lists 'Nutrition' (Eating an apple a day provides essential vitamins and minerals to the body.), 'Weight Loss' (Apples are low in calories and high in fiber, making them a great snack for weight loss.), and 'Cost-Effective' (Apples are generally inexpensive and are a cost-effective way to get the necessary nutrition.). The 'Cons' column lists 'Allergy Risk' (Some people may be allergic to apples and should avoid eating them.), 'Diabetes Risk' (Eating too many apples may increase blood sugar levels, which can be dangerous for those with diabetes.), and 'Stomach Upset' (Eating too many apples can cause stomach upset for some people.). A hand cursor icon is visible over the 'Allergy Risk' section. At the bottom right is the logo and name of '中信建投证券 CHINA SECURITIES'.

目录

一、ChatGPT：从全民关注的爆款应用走向商业化开放，在多个场景中展现价值

二、GPT4：更强大的多模态大模型，带来更广泛的商业应用可能

2.1.1 GPT-4: OpenAI最强大的多模态大模型

- 2023年3月15日，OpenAI发布多模态预训练大模型GPT-4，并称其是目前最强大的多模态大模型。其主要特点如下：
- **GPT-4具备从图像中获取和分析信息的能力：**GPT-4不仅拥有图像阅读能力，还可以从图像中提取逻辑，进行分析思考。以左图为例：问：如果把图中的绳子剪掉会怎么办？GPT-4: 气球会飞走。

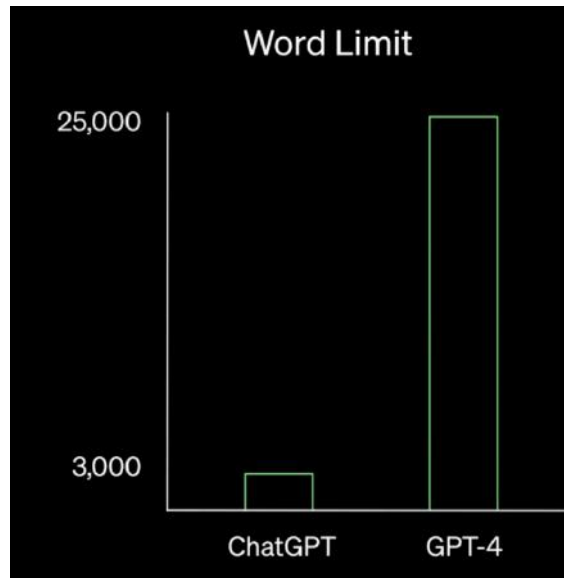
GPT-4对图片的理解能力、分析逻辑能力、物理等学科的知识储备都赋予了这个模型强大的智能。

- **GPT-4能够处理更长的文本输入：**在ChatGPT模型中，单次输入的文本量为3000个单词左右，而GPT-4模型单次能够处理超过25000个单词，是上一代ChatGPT模型的近8倍，代表着对长文章的阅读、分析、生成能力大幅提高。

图：GPT-4拥有从图像中提取逻辑的强理解力



图：GPT-4和ChatGPT单次输入的文本量对比



2.1.2 GPT-4：能力更为强大而全面的多模态预训练大模型

- **GPT-4更具备创造力和协作能力：**它可以与用户一起生成、编辑和迭代创意和技术写作任务，例如创作歌曲、编写剧本或学习用户的写作风格。如下左图所示，GPT-4根据用户要求生成了一个解释灰姑娘故事情节的句子，并且每个单词首字母需满足字母表A-Z的排序。
- **GPT-4相较于ChatGPT有更强的高级推理能力：**GPT-4能够回答出ChatGPT所不能回答的更为复杂的逻辑问题。如下右图所示，给定三个人物不同的空闲时间，要求模型安排出合适的会议时间，ChatGPT无法在此问题上给出正确答案（右图灰白色文字），而GPT-4能够合理的完成这一要求（右图绿色文字）。而在GPT-3.5难以处理的如AP Calculus BC（AP微积分）考试中，GPT-4的水平超过40%的应试者。

图：GPT-4根据用户要求创作

图：GPT-4能够处理ChatGPT无法处理的复杂推理问题

Input

Explain the plot of Cinderella in a sentence where each word has to begin with the next letter in the alphabet from A to Z, without repeating any letters.

要求用一句话描述灰姑娘的剧情且句子中每个单子的首字母需要满足字母表A-Z的排序

Output

A beautiful Cinderella, dwelling eagerly, finally gains happiness; inspiring jealous kin, love magically nurtures opulent prince; quietly rescues, slipper triumphs, uniting very wondrously, xenial youth zealously.

ChatGPT

Input

Andrew is free from 11 am to 3 pm, Joanne is free from noon to 2 pm and then 3:30 pm to 5 pm. Hannah is available at noon for half an hour, and then 4 pm to 6 pm. What are some options for start times for a 30 minute meeting for Andrew, Hannah, and Joanne?

Output

Here is a list of their availability:

- Andrew: 11 am to 3 pm
- Joanne: noon to 2 pm, and 3:30 pm to 5 pm
- Hannah: noon to 12:30 pm, and 4 pm to 6 pm

Based on their availability, there is a 30-minute window where all three of them are available, which is from 4 pm to 4:30 pm. So, the meeting can be scheduled at 4 pm.

GPT-4

Input

Andrew is free from 11 am to 3 pm, Joanne is free from noon to 2 pm and then 3:30 pm to 5 pm. Hannah is available at noon for half an hour, and then 4 pm to 6 pm. What are some options for start times for a 30 minute meeting for Andrew, Hannah, and Joanne?

Output

Andrew: 11 am - 3 pm
Joanne: 12 pm - 2 pm, 3:30 pm - 5 pm
Hannah: 12 pm - 12:30 pm, 4 pm - 6 pm

Common availability for a 30-minute meeting: 12 pm - 12:30 pm

ChatGPT生成的答案是错误的而GPT-4给出了正确的答案

2.2.1 GPT-4: 从语言模型走向多模态模型, 带来更为丰富的应用场景

- GPT-4是大型多模态模型, 能够接受图像和文本两种模态的输入, 再通过文字给出正确的文本答复, 在多个不同的应用场景中展现出强大的能力。

1) 给出一道物理题和相应的示意图, GPT-4可以给出具体的解题步骤和相应的答案(如下图1所示)

2) 给出论文的截图, GPT-4可以总结出论文的主要内容, 并根据用户进一步的提问补充其中的细节, 在用户的指令下, GPT-4进一步解释了论文中的第二个图像涉及的具体步骤。(如下图2所示)

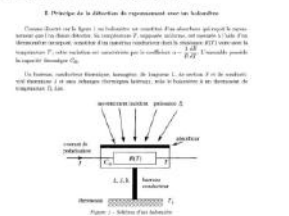
图1: GPT-4处理图像相关的物理问题

图2: GPT-4根据论文截图给出摘要(左)并回答相关问题(右)

用户输入物理题
及其对应示意图

GPT-4 visual input example, Solving École Polytechnique's Physics Exam:

User: Answer question I.1.a. Think step-by-step.



GPT-4

To answer question I.1.a, we need to express the temperature $T(x)$ in every point of the conductive bar, which is characterized by the abscissa x . We are given that $T(0) = T_0$ and $T(L) = T_b$.

For a stationary regime, the heat equation for the one-dimensional heat conduction in the bar is:

$$d^2T/dx^2 = 0$$

Integrating once, we get:

$$dT/dx = C_1$$

Integrating a second time:

$$T(x) = C_1x + C_2$$

To determine the constants C_1 and C_2 , we use the boundary conditions:

$T(0) = T_0$, which gives $C_2 = T_0$

$T(L) = T_b$, which gives $C_1 = (T_b - T_0)/L$

Finally, the expression for $T(x)$ is:

$$T(x) = (T_b - T_0) \cdot (x/L) + T_0$$

This represents a linear (temperature) profile along the conductive bar, with a slope of $(T_b - T_0)/L$.

GPT-4分步骤
给出解答


用户输入
论文图像
及摘要指令

GPT-4

GPT-4生成
论文摘要

GPT-4 visual input example, Pixel to Paper Summaries:

User: Below is part of the InstructGPT paper. Could you read and summarize it to me?



The InstructGPT paper focuses on training large language models to follow instructions with human feedback. The authors note that making language models larger doesn't inherently make them better at following a user's intent. Large models can generate outputs that are untruthful, toxic, or simply unhelpful.

To address this issue, the authors fine-tune language models on a wide range of tasks using human feedback. They start with a set of labeler-written prompts and responses, then collect a dataset of labeler demonstrations of the desired model behavior. They fine-tune GPT-3 using supervised learning and then use reinforcement learning from human feedback to further fine-tune the model. The resulting model, called InstructGPT, shows improvements in truthfulness and reductions in toxic output generation while having minimal performance regressions on public NLP datasets.

The authors conclude that fine-tuning with human feedback is a promising direction for aligning language models with human intent.

User:

GPT-4:

Could you explain the process described in Figure 2?

Figure 2 in the InstructGPT paper illustrates the three steps of their method:

1. **Supervised fine-tuning (SFT):** They collect demonstration data, which consists of labeler-written prompts and desired output behavior. This data is used to fine-tune GPT-3 with supervised learning.
 2. **Reward model (RM) training:** They collect comparison data, where a labeler ranks the outputs from several model samples (from best to worst). This data is used to train the reward model.
 3. **Reinforcement learning using Proximal Policy Optimization (PPO):** They optimize a policy against the reward model using reinforcement learning. For each new prompt sampled from the dataset, the policy generates an output. The reward model calculates a reward for the output, and the reward is used to update the policy using the PPO algorithm.
- These three steps are iteratively performed to train the InstructGPT model, which aims to generate outputs that align better with human preferences and follow instructions more effectively.

GPT-4根据用户要求
补充解释论文中图像的具体步骤

2.2.2 GPT-4: 从语言模型走向多模态模型, 带来更为丰富的应用场景

3) 用户给出一张具有相关数据信息的图表并提出了一个和图表数据相关的问题, GPT-4可以识别出图表中对应的数据, 给出问题的处理步骤和处理结果。如下图3所示, 给定一张含有芬兰、格鲁吉亚、西亚每日人均肉类消费的图表, GPT4按照要求计算出了格鲁吉亚和西亚每日人均肉类消费的和并给出了对应步骤。

4) 输入一张图片并提出和图片相关的问题, GPT-4可以基于图片给出准确的回答。如下图4所示, 输入一张“VGA线充手机”的图片并要求GPT-4解释图片幽默的原因, GPT-4成功按照用户的要求给出了图片的三个部分的具体含义并正确给出了图片之所以看上去滑稽是因为“将较大的VGA数据线接到了相对迷你的现代手机上”。而在图5的例子中, GPT-4成功解释了关于神经网络层数的漫画内容。

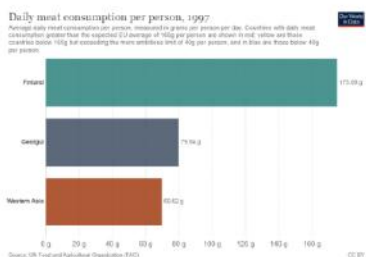
图3: GPT-4提取图像中的数据并完成相应计算

图4: GPT-4解释图片含义及幽默的原因

图5: GPT-4解释漫画的内容

GPT-4 visual input example, Chart Reasoning:

User: What is the sum of average daily meat consumption for Georgia and Western Asia? Provide a step-by-step reasoning before providing your answer.



GPT-4: To find the sum of average daily meat consumption for Georgia and Western Asia, we will follow these steps:

1. Identify the average daily meat consumption for Georgia.
 2. Identify the average daily meat consumption for Western Asia.
 3. Add the values from steps 1 and 2.
- Step 1: From the provided chart, we see that the average daily meat consumption for Georgia is 79.84 grams per person per day.
- Step 2: From the chart, we see that the average daily meat consumption for Western Asia is 69.62 grams per person per day.
- Step 3: Add the values from steps 1 and 2.
- $79.84 \text{ g (Georgia)} + 69.62 \text{ g (Western Asia)} = 149.46 \text{ g}$
- So, the sum of average daily meat consumption for Georgia and Western Asia is 149.46 grams per person per day.

Example of GPT-4 visual input:

User: What is funny about this image? Describe it panel by panel.

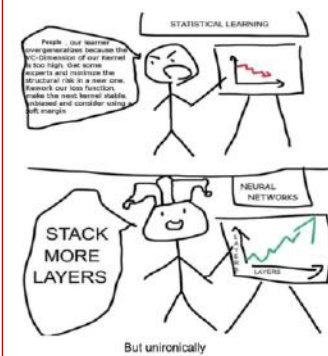


Source: <https://www.reddit.com/r/forums/comments/ub4bnr/zoom/>

- The image shows a package for a "Lightning Cable" adapter with three panels.
- Panel 1: A smartphone with a VGA connector (a large, blue, 15-pin connector typically used for computer monitors) plugged into its charging port.
- Panel 2: The package for the "Lightning Cable" adapter with a picture of a VGA connector on it.
- Panel 3: A close-up of the VGA connector with a small Lightning connector (used for charging iPhones and other Apple devices) at the end.
- The humor in this image comes from the absurdity of plugging a large, outdated VGA connector into a small, modern smartphone charging port.

GPT-4 visual input example, Moar Layers:

User: Can you explain why this is funny. Think about it step-by-step.



- GPT-4: The comic is satirizing the difference in approaches to improving model performance between statistical learning and neural networks.
- In statistical learning, the character is shown to be concerned with overfitting and suggests a series of complex and technical solutions, such as minimizing structural risk, reworking the loss function, and using a soft margin.
- In contrast, the neural networks character simply suggests adding more layers to the model. This is often seen as a common solution to improving performance in neural networks, but it's also considered a simplistic and brute-force approach.
- The humor comes from the contrast between the complexity and specificity of the statistical learning approach and the simplicity and generality of the neural network approach. The "But unironically" comment adds to the humor by implying that, despite being simplistic, the "stack more layers" approach is often effective in practice.

2.3 GPT-4: 从“百科全书”逐步成为文理通吃的“专家”

- 相较于过去的GPT系列模型，GPT-4在更多应用领域成为专家，包括为机器学习模型评判标准和为人类设计的专业测试。
- 在为机器学习模型设计的传统基准上，GPT-4的表现大大优于了目前的最好的语言模型，以及大多数SOTA模型

相较于GPT-3.5以及当前最好的语言模型，GPT-4在机器学习模型学术基准数据集上的表现实现了全面突破，在7项数据集上的测试均优于当前最优的语言模型。除了在Drop基准数据集上与QDGAT模型有一定差距，GPT-4在其余6项基准数据集的测试都实现了当前世界最优秀的性能。在MMLU数据集中，GPT-4不仅在英语语言上的性能实现了突破，在其他语言方面的表现也优于现在的语言模型。

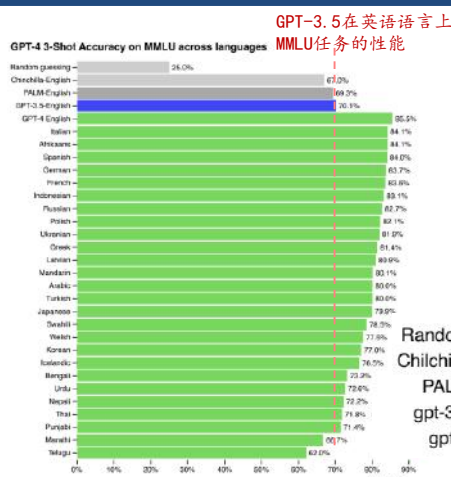
- 在为人类设计的各种专业测试和学术基准数据集中，GPT-4展现出媲美甚至优于人类的水平。

GPT-3.5模型在一些相对简单的人类专业测试中已经展现出不错的性能，但当任务的复杂度上升后，其便难以处理，而GPT-4模型在一些相对困难的任务中也已经取得相当优异的性能，例如其在模拟律师考试中取得了应试者中前10%的成绩，而GPT-3.5模型只有倒数10%的成绩，甚至在GPT-3.5难以处理的如AP Calculus BC（AP微积分）考试中，GPT-4的水平也已经超过40%的应试者。

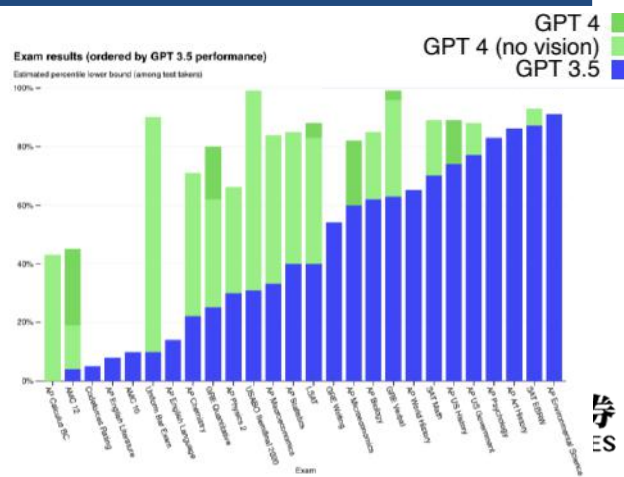
表：GPT-4在机器学习模型学术基准数据集上的表现

	GPT-4 Evaluated few-shot	GPT-3.5 Evaluated few-shot	语言模型的 最佳性能 LM SOTA Best external LM evaluated few-shot	所有模型的 最佳性能 SOTA Best external model (incl benchmark-specific tuning)
MMLU [43] Multiple-choice questions in 57 subjects (professional & academic)	86.4% 5-shot	70.0% 5-shot	70.7% 5-shot U-PaLM [44]	75.2% 5-shot PaLM [45]
HellaSwag [46] Commonsense reasoning around everyday events	95.3% 10-shot	85.5% 10-shot	84.2% LLaMA (validation set) [24]	85.6% ALUM [47]
A12 Reasoning Challenge (ARC) [48] Grade-school multiple-choice science questions. Challenge set.	96.3% 25-shot	85.2% 25-shot	85.2% 8-shot PaLM [49]	86.5% ST-MOE [18]
WinoGrande [50] Commonsense reasoning around person resolution	87.5% 5-shot	81.6% 5-shot	85.1% 5-shot PaLM [3]	85.1% 5-shot PaLM [3]
HumanEval [37] Python coding tasks	67.0% 0-shot	48.1% 0-shot	26.2% 0-shot PaLM [3]	65.8% CodeT5 + GPT-3.5 [51]
DROP [52] (F1 score) Reading comprehension & arithmetic.	80.9 3-shot	64.1 3-shot	70.8 1-shot PaLM [3]	88.4 QDGAT [53]
GSM-8K [54] Grade-school mathematics questions	92.0%* 5-shot chain-of-thought	57.1% 5-shot	58.8% 8-shot Mixtral [55]	87.3% ChiChiLLa + SFT + GORM-R ORM re-ranking [56]

图：GPT-4在MMLU不同语言上的性能



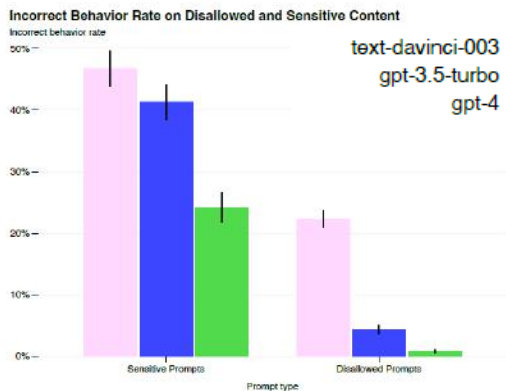
图：GPT-4在人类专业测试中的表现



2.4 GPT-4：在可控性和真实性方面比ChatGPT有大幅提升

- **GPT-4更加可控，回答更加符合事实：**Open AI 花费了 6 个月时间使用对抗性测试程序和ChatGPT的经验教训对GPT-4进行迭代调整，从而在可控性和真实性上取得了很好的结果。在内部评估中，GPT-4 回答可能生成有害内容的输入的可能性降低了 82%，产生符合事实回复的可能性高出 40%。
- **引入人类反馈进行训练。**Open AI引入了ChatGPT用户以及50名来自人工智能、网络安全、生物安全和国际安全等领域的专家提交的反馈来训练模型，以使模型的回答避免含有某些不恰当的言论。
- **从实际使用中不断提升。**Open AI将以前模型的使用经验应用到 GPT-4 的安全研究和监控系统中。例如Open AI会自动检测用户的行为是否违背安全性原则。若违反，系统会临时发布警告暂停或禁止用户的行为。与 ChatGPT 一样，随着越来越多的人使用它，Open AI将定期更新和改进 GPT-4。
- **GPT-4协助进行安全方面的研究。**GPT-4 的高级推理和指令遵循功能加快了安全研究工作。Open AI使用 GPT-4 来帮助创建训练数据，以便在训练、评估和监控中对分类器进行微调和迭代。

图：GPT-4降低危险行为比率



资料来源：Open AI，中信建投

图：Open AI通过三方面提升GPT-4可控性和真实性

引入人类的反馈进行训练

从实际使用中持续提升

GPT-4协助进行安全方面的研究

Training with human feedback

We incorporated more human feedback, including feedback submitted by ChatGPT users, to improve GPT-4's behavior. We also worked with over 50 experts for early feedback in domains including AI safety and security.

Continuous improvement from real-world use

We've applied lessons from real-world use of our previous models into GPT-4's safety research and monitoring system. Like ChatGPT, we'll be updating and improving GPT-4 at a regular cadence as more people use it.

GPT-4-assisted safety research

GPT-4's advanced reasoning and instruction-following capabilities expedited our safety work. We used GPT-4 to help create training data for model fine-tuning and iterate on classifiers across training, evaluations, and monitoring.

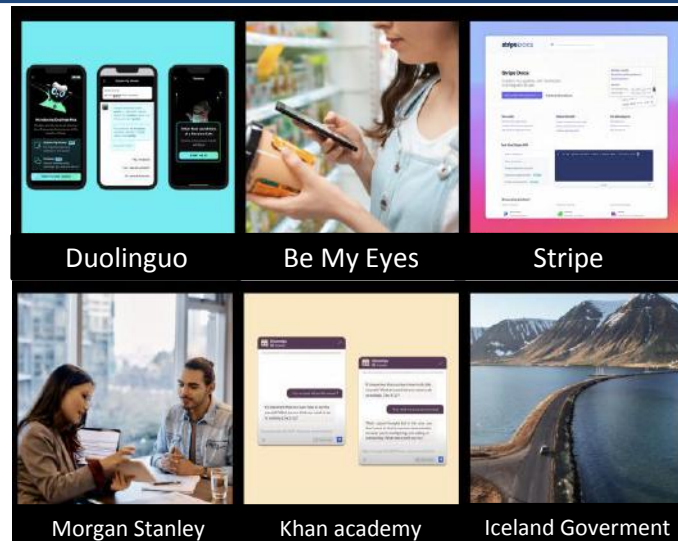
2.5 ChatGPT升级GPT-4版本，GPT-4开放相应API加速商业化，

- **ChatGPT升级了对应的GPT-4版本：**在GPT-4发布后，OpenAI直接将GPT-4接入了ChatGPT。订阅了ChatGPT Plus的用户可以有使用上限的GPT-4访问权限。
- **GPT-4开放了相应的API：**GPT-4纯文本API可以通过注册申请的方式获取，但目前图像输入仍处于Alpha测试阶段，尚未开放给用户。价格方面，输入长度为8K的版本输入和回复的价格分别为每1000个token 0.03美金和0.06美金，输入长度为32K的版本输入和回复的价格分别为每1000个token 0.06美金和0.12美金。
- **GPT-4商业化落地加速：**GPT-4强大的性能和解决丰富场景下具体问题的能力让其在优化人机交互体验、提供专业服务、提升组织效能、文化传承与保护等方面都展现了巨大的潜能。

表：GPT-4和ChatGPT API 价格对比

图：GPT-4部分的落地应用案例

模型名	定价	输入长度
GPT-3.5-turbo	0.002美金/ 1K tokens	
	0.03美金/ 1K tokens	8K长度的输入
GPT-4（纯文本）	0.06美金/ 1K tokens	8K长度的回复
	0.06美金/ 1K tokens	32K长度的输入
	0.12美金/ 1K tokens	32K长度的回复



2.5.1 GPT-4商业应用：工作方式的革新

- **Microsoft 365 Copilot：结合GPT-4连通多种应用**

Microsoft 365 Copilot将GPT-4的强大功能与 Microsoft Graph 和 Microsoft 365 应用中的数据相结合，通过自然语言输入即可产生强大生产力，极大提升了工作效率。

个人办公：Copilot被整合进所有的Microsoft 365应用。在Word中，用户简单描述即可让其生成初稿，也可输入其它文件，指定内容进行创作。在Excel中，它可以分析趋势并进行数据可视化，用户不再需要使用复杂的公式；在PPT中，它可以根据用户的要求生成演示文稿，也可将其它文件转为精美PPT；在Outlook中，它可以挑选出有价值的邮件并总结对应要点，还可以提供回复建议。

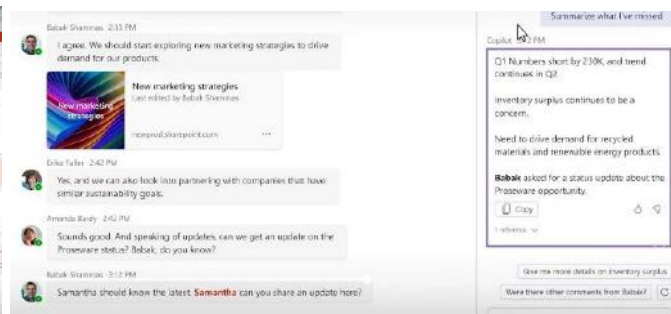
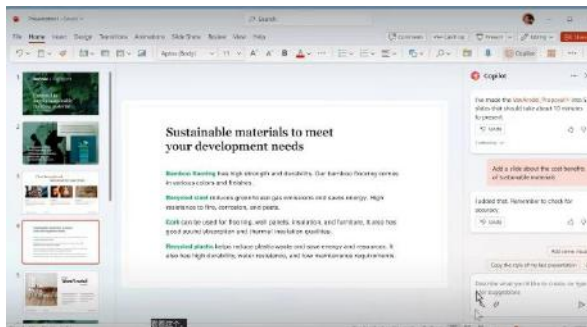
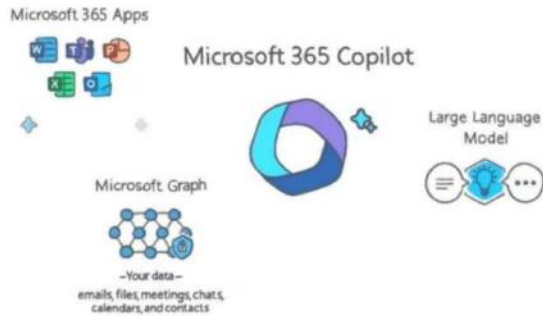
团队协作：Copilot不仅局限于Office的几个软件，还将整个办公生态全部打通。邮件、联系人、在线会议等等软件的数据都被接入大语言模型，构成新的Copilot系统。Copilot会将错过的会议消息汇总成报告。在会议中它能够实时做出总结。会议结束，Copilot还可根据内容自动生成跟进邮件。此外，它还可以将重复的工作流程自动化，为公司做SWOT分析，实时协同工作等。

- 我们认为，结合了GPT-4强大AI能力的Microsoft 365 Copilot，将带来工作效率的大幅提升和工作方式的革新。这一强大增值服务的落地将带动Office ARPU值（每个用户贡献收入平均值）的提升，并推动Office软件整体渗透率的提升，具有广阔的市场前景。若按Microsoft 365 Office2022年用户数3.45亿人，保守参考GitHub Copilot 10美金/月的订阅费，假设其中20%的用户升级相关服务来计算，每年大约将带来超80亿美元的收入。

图：Copilot系统

图：Copilot基于简单描述生成PPT

图：Copilot总结错过的会议内容



2.5.2 GPT-4商业应用：人机交互体验优化

• Duolingo：使用GPT-4改进对话功能

Duolingo是一个语言学习平台，每月有超过50万用户使用其进行外语学习，提供40多种语言课程，支持100多种语言。与GPT-4合作的产品Duolingo Max产品目前有两项功能：一是Roleplay，即角色扮演，用户可以在众多的场景中和AI伙伴进行对话练习，如下左图便是一个模拟用法语在咖啡厅场景中点单的例子；二是Explain My Answer，即答案解释，会针对用户的回答对上下文的错误进行反馈，若用户对解释尚不满意，可以要求获得进一步的解释。目前Duolingo Max应用于西班牙语和法语，未来将应用于更多语言，并开发更多功能。

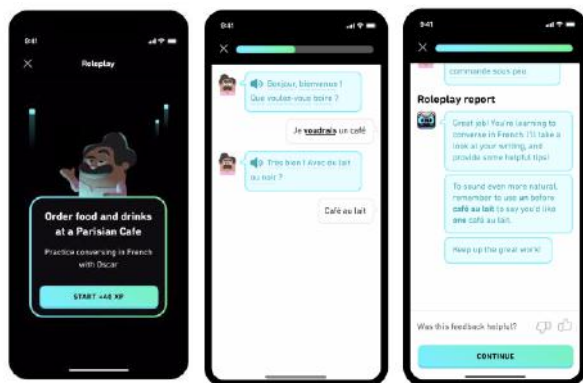
• Be My Eyes：使用GPT-4优化视觉辅助功能

Be My Eyes是一家创立于2012年的丹麦公司，持续为超过2.5亿视力障碍或弱视人群提供辅助技术。借助GPT-4的多模态技术，Be My Eyes在其应用程序中正在开发Virtual Volunteer功能，其可以产生与人类志愿者相同水平的分析、理解和反馈能力。例如下图所示，对于两件条纹衬衫进行拍照，便可以回复与图片相关的问题，如“哪一件是红色条纹衬衫”、“红色条纹衬衫之外的另外一件是什么样的”。Be My eyes的首席执行官表示，这项技术将满足视力障碍或弱视人群大量的视觉解释需求，使他们在生活中拥有更大的独立性。

图：Duolingo max的roleplay功能

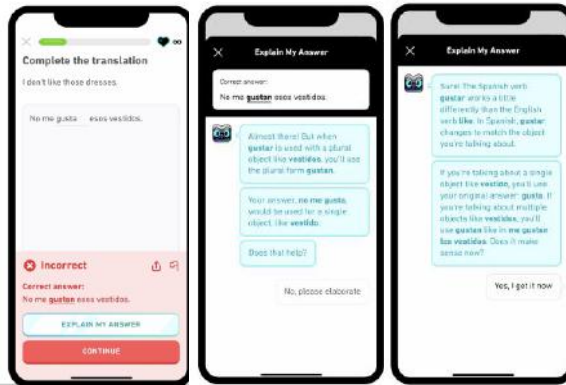
图：Duolingo max的explain my answer功能

图：Virtual Volunteer基于拍照图片回答相关问题



模拟在咖啡厅中用法语进行点单并给出相应评价

资料来源：OpenAI，中信建投



针对用户的回答对具体知识进行解释，
并可以根据用户要求进一步扩充相关知识



基于用户输入的照片给出相关问题的答案

2.5.3 GPT-4商业应用：提供专业且个性化的服务

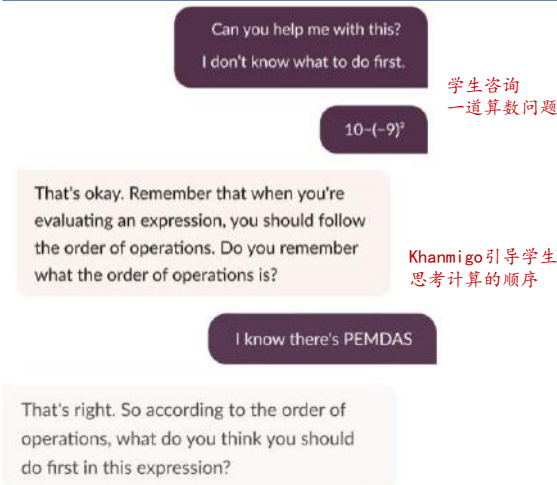
• Khan Academy：结合GPT-4的教育平台

Khan Academy是一个非盈利的教育平台。其2022年开始测试最新版本的 Open AI 语言模型，将向部分参与者提供 Khanmigo 试点计划。结合GPT-4打造的 Khanmigo AI助手可以为学习者提供定制化的教育以及为教师提升工作效率。

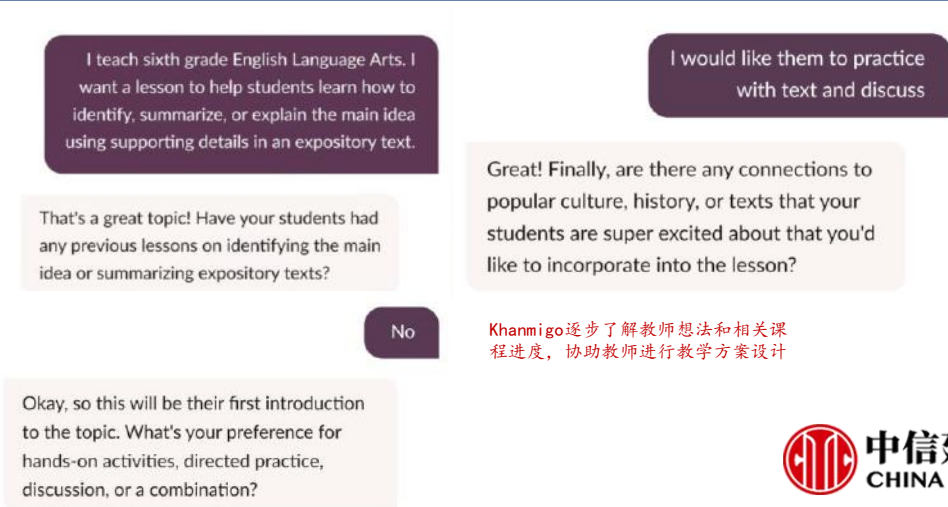
为学习者提供定制化教育：GPT-4像人类一样多轮对话的能力使得 Khan Academy 能够向每个学生提出个性化的问题，以促使其更深入地学习。如下左图所示，学习者向 Khanmigo 询问数学题时，Khanmigo没有直接给出答案，而是通过问题的形式一步步引导学习者回忆学过的知识，并提供解决问题的思路。这样能够使学习者更加深入地思考，达到更好学习效果。

为教师提高工作效率：教师可以利用Khanmigo编写课堂提示或为课程创建教学材料。如下右图所示，Khanmigo通过问答逐步了解教师想法和课程进度，而后为教师提供教学设计的方案。

图：Khanmigo一步步引导学生回答问题



图：Khanmigo帮助教师设计教学方案



2.5.4 GPT-4商业应用：提升组织运行效能

• Stripe：结合GPT-4的支付平台

Stripe是一个支付平台，借助GPT4的能力进行更好地服务客户、回答技术问题以及识别虚假信息。

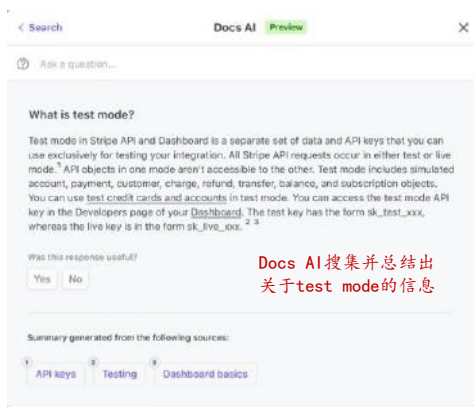
在客户服务方面，Stripe发现部分客户会隐藏它们的网站，因此在跟进服务时搜索它们需要花费很多时间精力。借助GPT-4，Stripe能够扫描这些网站并且做出汇总，相比于人工节省了很多精力。

在回答技术问题方面，借助GPT-4，Stripe能够快速理解用户的问题，代替他们阅读详细的文档，梳理想关信息并总结出一个解决方案，这将极大地提升工作效率。

在识别虚假信息方面，在Stripe支持的社区Discord中有很多不怀好意的使用者是为了获取用户信息或者散步虚假信息才加入其中。GPT-4 通过分析 Discord 中帖子的语法，就会发现欺诈团队，以方便相关工作人员将他们清除出去。

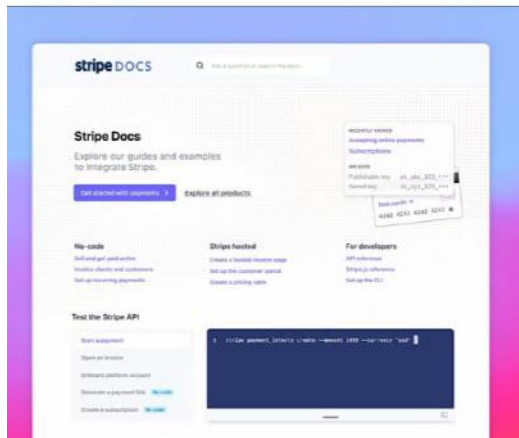
此前，Stripe也曾整合了GPT-3以更好服务客户，包括总结客户问题的功能。

图：利用Stripe中的Docs AI搜集信息



Docs AI搜集并总结出
关于test mode的信息

图：Stripe引入GPT-4去提升用户体验和对抗虚假信息



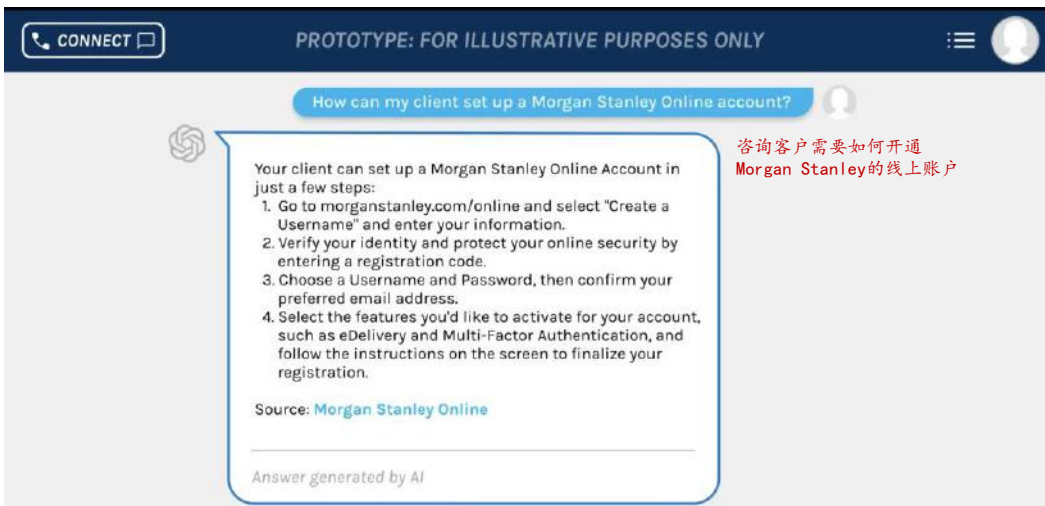
2.5.5 GPT-4商业应用：提升组织运行效能

• Morgan Stanley：利用GPT-4来实现企业运作中多维度的提升

与其他公司使用GPT系列模型来提升工作效率相似，GPT-4模型从海量的预训练数据集中学到的海量知识以及高效的处理文本间信息的能力，也被Morgan Stanley用于协助员工实现工作效率的提升。如下图便是通过咨询系统帮助解答客户如何开通线上账户的问题。目前每天有超过200名员工在系统上进行相关问题的咨询并得到他们需要的反馈。

Morgan Stanley作为成立时间近百年的全球财富管理领导者，其内部数据库中包含了大量高质量的涵盖全球资本市场、资产类别、行业分析和经济区域等方面的见解。这些知识以PDF形式存储在多个内部网站中，搜索和扫描以找到特定问题的答案是一个相当繁琐的过程，而GPT-4的应用让这些丰富而深刻的知识以一种更可用和可操作的格式出现，助力公司内部数据与知识应用高效化。

图：咨询GPT-4如何在Morgan Stanley开通线上账户



资料来源：OpenAI，中信建投

2.5.6 GPT-4商业应用：文化保护与传承

● 冰岛政府：利用GPT4来实现对于语言的保护

在英语逐渐成为主流的工作语言的当下，使用人口仅三十余万的冰岛语面临着语言和文化层面的巨大冲击。在过去，由于大规模语言模型的训练数据往往来自互联网上的广泛语料，而这些语料大都是英语或其他使用人口更为广泛的语言，因此对于像冰岛语一样的小众语种效果较差。

GPT-4的出现改变了这一情况，一方面，其实现了相较于GPT-3.5而言更为优秀的对于冰岛语的理解和生成能力，另一方面，对于GPT-4的冰岛语能力的“培训”（实质是微调）也变得更为容易。因而目前GPT-4不但可以实现较为准确的冰岛语翻译（之前GPT-3和ChatGPT都还无法完成）还可以创造一首来自北欧神话的古代冰岛诗，甚至是根据冰岛当地的文化特点进行相应问题的回答，如右图所示，若提问者使用冰岛语进行提问，GPT-4则会基于冰岛的情况进行回答。在未来将有希望发展出基于冰岛语乃至其他小众语种专用的聊天机器人，这极大地推进了对冰岛语这一类小众语种的保护。

表：GPT-4实现了过去的GPT系列模型都无法实现的冰岛语翻译

	GPT-3	ChatGPT	GPT-4
Icelandic	Donald Duck heitir á íslensku Donaldi Kjáni	Donald Duck heitir sama nafni á íslensku og ensku	Donald Duck heitir Andrés Önd á íslensku
English	Donald Duck is called in Icelandic Donaldi the Fool	Donald Duck has the same name in Icelandic and English	Donald Duck is called Andrés Önd in Icelandic
	Incorrect	Incorrect	Correct

表：GPT-4可以基于不同的输入语言给出对应的答案

Icelandic	English
Hvað sitja margir á þingi?	How many parliamentarians are there?
Á Alþingi, sem er lögþing Íslands, sitja 63 þingmenn. In Alþingi, which is the Icelandic parliament, there are 63 parliamentarians.	It depends on which country you are referring to. For example: – In the United Kingdom, the Parliament has 650 members in the House of Commons and approximately 800 members in the House of Lords. – In Canada, there are 338 members of Parliament in the House of Commons and 105 senators in the Senate. – In India, there are 545 members in the Lok Sabha (House of the People) and 245 members in the Rajya Sabha (Council of States). – In Australia, there are 151 members of the House of Representatives and 76 senators in the Senate.

风险提示

- **GPT技术发展不及预期：**GPT属于先进AI算法，若后续GPT算法更新迭代效果不及预期，则会影响GPT演进及拓展，进而会影响其商业化落地等；
- **算法歧视与人权道德风险：**GPT引入基于人工反馈的强化学习，在人工反馈中，不可避免的会参杂歧视问题，从而会将歧视带入后续的人机交互中；大模型在进行预训练过程中，使用了大量数据，其中会涉及相关隐私问题；
- **数据数量与数据质量不及预期：**大模型需要大量的高质量数据进行训练，若数据数量和质量存在短板，则会影响模型效果。
- **GPT接口开放不明确：**目前GPT是否能对国内开放API还不明朗，若未来无法接入，可能导致相关应用在国内无法落地。
- **算力受限：**目前大模型训练以GPU为主，尤其是英伟达的GPU。未来若GPU产能或者中美摩擦升级，可能导致国内厂商的算力布局。

感谢陈思玥对本报告的贡献。

分析师介绍

阎贵成：中信建投证券通信&计算机行业首席分析师，北京大学学士、硕士，专注于云计算、物联网、信息安全、信创与5G等领域研究。近8年中国移动工作经验，6年多证券研究经验。系2019-2021年《新财富》、《水晶球》通信行业最佳分析师第一名，2017-2018年《新财富》、《水晶球》通信行业最佳分析师第一名团队核心成员。

金戈：中信建投证券研究发展部计算机行业联席首席分析师，帝国理工学院工科硕士，擅长云计算、金融科技、人工智能等领域。

于芳博：中信建投计算机行业分析师，北京大学空间物理学学士、硕士，2019年7月加入中信建投，主要覆盖方向智能汽车、CPU/GPU/FPGA/ASIC、EDA和工业软件等方向。

评级说明

投资评级标准		评级	说明
报告中投资建议涉及的评级标准为报告发布日后6个月内的相对市场表现，也即报告发布日后的6个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。A股市场以沪深300指数作为基准；新三板市场以三板成指为基准；香港市场以恒生指数作为基准；美国市场以标普 500 指数为基准。	股票评级	买入	相对涨幅15%以上
		增持	相对涨幅5%—15%
		中性	相对涨幅-5%—5%之间
		减持	相对跌幅5%—15%
		卖出	相对跌幅15%以上
	行业评级	强于大市	相对涨幅10%以上
		中性	相对涨幅-10-10%之间
		弱于大市	相对跌幅10%以上



中信建投证券
CHINA SECURITIES

分析师声明

本报告署名分析师在此声明：(i) 以勤勉的职业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，结论不受任何第三方的授意或影响。(ii) 本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

法律主体说明

本报告由中信建投证券股份有限公司及/或其附属机构（以下合称“中信建投”）制作，由中信建投证券股份有限公司在中华人民共和国（仅为本报告目的，不包括香港、澳门、台湾）提供。中信建投证券股份有限公司具有中国证监会许可的投资咨询业务资格，本报告署名分析师所持中国证券业协会授予的证券投资咨询执业资格证书编号已披露在报告首页。

本报告由中信建投（国际）证券有限公司在香港提供。本报告作者所持香港证监会牌照的中央编号已披露在报告首页。

一般性声明

本报告由中信建投制作。发送本报告不构成任何合同或承诺的基础，不因接收者收到本报告而视其为中信建投客户。

本报告的信息均来源于中信建投认为可靠的公开资料，但中信建投对这些信息的准确性及完整性不作任何保证。本报告所载观点、评估和预测仅反映本报告出具日该分析师的判断，该等观点、评估和预测可能在不发出通知的情况下有所变更，亦有可能因使用不同假设和标准或者采用不同分析方法而与中信建投其他部门、人员口头或书面表达的意见不同或相反。本报告所引证券或其他金融工具的过往业绩不代表其未来表现。报告中所含任何具有预测性质的内容皆基于相应的假设条件，而任何假设条件都可能随时发生变化并影响实际投资收益。中信建投不承诺、不保证本报告所含具有预测性质的内容必然得以实现。

本报告内容的全部或部分均不构成投资建议。本报告所包含的观点、建议并未考虑报告接收人在财务状况、投资目的、风险偏好等方面的具体情况，报告接收者应当独立评估本报告所含信息，基于自身投资目标、需求、市场机会、风险及其他因素自主做出决策并自行承担投资风险。中信建投建议所有投资者应就任何潜在投资向其税务、会计或法律顾问咨询。不论报告接收者是否根据本报告做出投资决策，中信建投都不对该等投资决策提供任何形式的担保，亦不以任何形式分享投资收益或者分担投资损失。中信建投不对使用本报告所产生的任何直接或间接损失承担责任。

在法律法规及监管规定允许的范围内，中信建投可能持有并交易本报告中所述公司的股份或其他财产权益，也可能在过去12个月、目前或者将来为本报告中所述公司提供或者争取为其提供投资银行、做市交易、财务顾问或其他金融服务。本报告内容真实、准确、完整地反映了署名分析师的观点，分析师的薪酬无论过去、现在或未来都不会直接或间接与其所撰写报告中的具体观点相联系，分析师亦不会因撰写本报告而获取不当利益。

本报告为中信建投所有。未经中信建投事先书面许可，任何机构和/或个人不得以任何形式转发、翻版、复制、发布或引用本报告全部或部分内容，亦不得从未经中信建投书面授权的任何机构、个人或其运营的媒体平台接收、翻版、复制或引用本报告全部或部分内容。版权所有，违者必究。

中信建投证券研究发展部

北京
东城区朝内大街2号凯恒中心B
座12层
电话：(8610) 8513-0588
联系人：李祉瑶
邮箱：lizhiyao@csc.com.cn

上海
浦东新区浦东南路528号南塔2106
室
电话：(8621) 6882-1612
联系人：翁起帆
邮箱：wengqifan@csc.com.cn

深圳
福田区益田路6003号荣超商务中心
B座22层
电话：(86755) 8252-1369
联系人：曹莹
邮箱：caoying@csc.com.cn

中信建投（国际）

香港
中环交易广场2期18楼
电话：(852) 3465-5600
联系人：刘泓麟
邮箱：charleneliu@csc.hk