

考试碾压人类90%“做题家” 可识别图片 还会角色扮演 GPT-4来了

美国当地时间3月14日,OpenAI官网发布了GPT-4,一石激起千层浪。

比起此前上亿人争相试用的ChatGPT(依靠GPT-3.5模型),GPT-4在日常对话表现中并没有太大革新性的进步。但作为一个“多模态”模型,GPT-4不仅适用更多语种,而且在语言文本之外新增了强大的识图功能。

在对话中,GPT-4比ChatGPT更有创造性和推理能力,不仅可以创作诗歌、帮你安排工作日程,还在模拟的标准化考试中碾压绝大多数人类考生。令人“细思极恐”的是,和以前人们印象中“不近人情”的AI相比,GPT-4在图片和文字领域都更“善解人意”,甚至能扮演不同的人格和用户对话。



会创作诗词,还能押上韵

ChatGPT的记忆长度大约是8000个单词,折合四到五页书。GPT-4的最大记忆约6.4万个单词,折合近50页文字,因此对话具有更强的准确度和上下文连贯性。

与ChatGPT相比,GPT-4在文本对话中最大的进步在于具备了一定的文学创作能力,能根据

要求创造出不同文风的歌词、诗篇甚至短句。更夸张的是,它甚至能押上韵脚。OpenAI官网展示了一段另类押韵的炫技文本。

根据指令,GPT-4用26个单词写成一句话概括“灰姑娘”的童话故事。令人震撼的是,这26个单词的首字母恰好是从A到Z的排序。

可识别图片,还能get到搞笑的梗



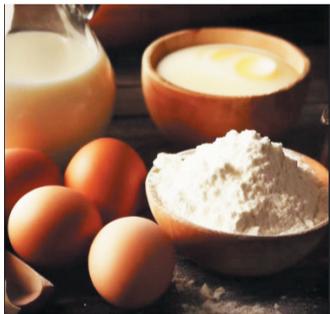
和单纯文本对话的ChatGPT相比,GPT-4最突出的地方在于可以接受图片输入,它会处理这些图像以查找相关信息。GPT-4不仅能简单识别图片要素并进行描述,甚至具备读懂梗图的能力。

OpenAI提供了一个经典示例,输入一张超大的图片,模型能解读出这张图片搞笑的点在哪里:

图片显示了一个“Lightning Cable”适配器的包装,总共包含三个部分。图1:一部智能手机,其充电端口插着一个VGA接口(大的蓝色15针接口,通常用于电脑显示器)。图2:“Lightning Cable”适配器的包装上有一张VGA接口的图片。图3:VGA连接器的特写,末端是一个小的Lightning连接器(用于为iPhone和其他苹果设备充电)。这张图片的笑点来自将一个巨大、过时的VGA连接器插入一个小体积、现代的智能手机充电端口的荒谬性。

此外,GPT-4还和为视力障碍者服务的公益项目Be My Eyes合作。AI模型可以为盲人或视弱群体描述裙子上的图案,帮助识别植物,解释如何使用健身房的某台机器。

会推理判断,考试能力超过“做题家”



OpenAI用上面这张图片演示了GPT-4的推理和识图能力。当模型收到这张在厨房里拍的照片后,它不仅能辨认出面粉、

鸡蛋等元素,还可以迅速给出食谱建议。

由于有着强大的推理能力,在标准化的技能测试中,GPT-4比起ChatGPT的表现有了实质性的提升。OpenAI官网称,在模拟美国律师从业资格考试中,ChatGPT的成绩在全体考生中只能排到末尾10%,而GPT-4能排进前10%。在难度更高的国际生物学联赛中,GPT-4甚至能吊打99%的人类选手。

在“美国高考”SAT中,GPT-4在数学和阅读部分也能战胜近90%的考生。

会“角色扮演”不同人格

在情商方面,GPT-4比起GPT-3.5也提高很多。根据OpenAI的描述,用户能够将ChatGPT原本固定冗长的语气和风格,改为更适合他们需要的个性化风格。

具体来说,用户可以要求模型假扮各种角色,比如电视台进

行采访的主持人、面对老年用户的基金经理等。根据案例展示,GPT-4可以熟练切换语言风格适应自己的角色定位。比如在和老年人对话时,使用平和缓慢的语气,而在担任经理下达命令时又采取果断坚决的态度。

■ 相关新闻

国内企业涌入赛道

美团联合创始人王慧文对于AI的态度很明确,“关于人工智能,到现在为止,我只有一个判断,必须参与。”

今年2月,王慧文在朋友圈发布“组队帖”,要带资5000万美元进入AI赛道,设立光年之外科技有限公司,资金占股25%,其余75%的股份用于邀请顶级研发人员,打造中国Open AI。

据王慧文透露,光年之外估值达2亿美元,下轮融资已有顶级VC认购2.3亿美元。3月8日,美团创始人王兴在朋友圈回应王慧文的宣言,宣布将以个人身份参与王慧文的创业公司“光年之外”的A轮融资,并出任董事。

在“双王”三度携手、进场AI赛道的同时,众多互联网大佬纷纷开始招兵买马,争相抢夺人才:原京东集团高级副总裁、集团技术委员会主席周伯文同样在朋友圈发布招募研发、产品及算法人员,原搜狗CEO王小川对媒体回应正快速筹备AI大模型,快手前AI核心人物李岩也投身“大模型”赛道创业,成立AI公司元石科技,进行多模态大模型的研发。

百度、腾讯、阿里、字节跳动等互联网大厂也在“争分夺秒”开发自有大模型。在OpenAI发布GPT4.0后,3月16日即将发布“文心一言”的百度压力再度增加。而腾讯针对类ChatGPT产品已成立“混元助手”项目组,阿里巴巴则表示其聊天机器人正处于研发内测阶段。

ChatGPT的背后离不开大模型、大数据、大算力,大厂想要追赶Open AI,也需要面临多方面的挑战。

在算力要求方面,市场调查机构TrendForce集邦咨询公布的最新报告指出,如果以英伟达A100显卡的处理能力计算,运行ChatGPT将需要使用到3万块英伟达GPU,目前英伟达A100显卡的售价在1万至1.5万美元之间,3万块售价超过3亿美金。

除算力要求高外,需要大量的数据集,成本非常高。其使用Transformer神经网络架构,训练数据来自包括维基百科,以及真实对话在内的庞大语料库,有着多达1750亿个模型参数。

■ 相关链接

专家点评GPT-4: 数学不太好的文科生

业内人士如何看待这个多模态预训练大语言模型?记者采访了上海交通大学教授、自然语言处理专家赵海。

“读图编程”表演令人惊艳

赵海教授解释,模态对应的是人类感知类型,包括文本、图像、声音等。一个智能系统如果能处理两种或两种以上的感知信息,就属于多模态。一些网络媒体对GPT-4有误解,以为它能根据用户输入的文字生成图片,其实它没有这个功能。GPT-4比ChatGPT多出的功能,是用户可以输入图片,它能“读图”并输出文字。

在执行编程任务上,这个多模态大模型也能“读图”。OpenAI的产品发布会上,研发人员演示了GPT-4能识别他手绘的一张网页草图,并根据草图写出网页的前端代码。虽然这个网页很简单,但它的“读图编程”还是令人惊艳,显示出较高的图片理解和编程能力。不过遗憾的是,这家美国企业表示,这项功能目前不对公众开放。

谈及这种多模态功能,赵海直言:“GPT-4的产品形态会让公众感到惊艳,但对一些业内人士来说,可能就没那么惊讶了。”在他看来,GPT-4强大功能的核心与ChatGPT一样,也是参数量高达1750亿的预训练大语言模型。至于图像分析能力,可看作给大语言模型“打补丁”、拼接图像向量的结果。

“从理论上说,给大语言模型拼接图像向量是不难的,但工程实现的难度很大。”赵海分析,“因为GPT-3模型已有千亿级参数,在此基础上增加大量图像参数,算力上的压力会非常大。”

这与达观数据首席执行官陈运文博士的观点颇为吻合。作为上海最大的自然语言模型研发企业创始人,陈运文认为,在大语言模型的工程基础领域,国内科技界和产业界的当务之急,是开发出高质量的大模型,在此基础上发展产业生态就会有很大潜力。这是“BAT”目前也

没有完全解决的问题。

数学解题能力靠“打补丁”

除了功能增加,GPT-4的文本处理和生成能力也优于ChatGPT。

谈及这个大模型的数学水平,赵海笑道:“很多人没考过GRE,其实GRE的数学题难度和中国小学数学题差不多。GPT-4的数学水平不会很高,很多小学奥数题可能就解不出来。”

GPT-4为何是“数学不太好的文科生”?他解释说,和图像分析能力一样,它的数学解题能力也是通过在大模型上“打补丁”才获得的。如果只接受过互联网上的文本大数据训练,GPT-4和ChatGPT只能做出简单的算术题。要让它解出难度高一点的代数和几何题,研发团队必须为大模型加入数学领域的解题算法。

“OpenAI说GPT-4有强大的推理能力,我感觉主要指的是它的数学水平高。”赵海告诉记者,“其实,推理能力包含逻辑推理能力和数学抽象思维能力,这两种能力是有区别的,前者侧重寻找因果关系。相对而言,GPT-4的逻辑推理能力更强。”

从OpenAI的GPT系列(GPT-1、2、3、ChatGPT和GPT-4)开发思路来看,这家美国企业在向着通用人工智能商用目标迈进。与ChatGPT相比,GPT-4离商用产品又进了一步。这种进步包括:在应对政治敏感、违反伦理或法律的指令和问题时,GPT-4表现得越发成熟和安全。一些体验过GPT-4的用户还反映,输入同样的问题,GPT-4的回答常常优于ChatGPT,表现为信息量更大,“车轱辘话”更少。

未来,GPT产品可能会成为一个类似电脑操作系统的平台,让通用人工智能进入千家万户。赵海指出,这种革命性产品的基础是预训练大语言模型,国内科技界和产业界的当务之急,是开发出高质量的大模型,在此基础上发展产业生态就会有很大潜力。