



# GPT-4:能读图解题讲冷笑话

## 比上一代更准确安全但仍存在偏见性和“幻觉”

### 比上一代强大在哪里

先来解释一下GPT,这是一种预训练语言模型,这种模型读得越多,就懂得越多。ChatGPT的核心即为GPT,只不过加上了能与人聊天的对话框。前阵子大火的ChatGPT用的是GPT-3.5,现在更新到了GPT-4。

GPT-4跟GPT-3.5具有相同的API接口和交互界面,但在文本总结和加工能力上,GPT-4有明显提升,表现在提出指令后,模型能更好地遵循指令给出答案。例如让一篇文章变成一个句子,每个单词都以字母“G”开头,GPT-3.5还未尝试就放弃了任务,而GPT-4可以很好地完成。

事实上,在许多方面,GPT-4能做到GPT-3.5力不能及的事情——

OpenAI让GPT-4参加了许多项人类通用考试,结果证明,它在许多测试和基准测试中的表现比前一代大为提高:GPT-4在俗称“美国高考”的SAT里分数增加了150分,能拿到1600分中的1410分;

它还能通过模拟律师考试,分数在应试者的前10%左右,相比之下,它的“前任”得分在倒数

中国版的ChatGPT“文心一言”今天就要问世了,但大洋彼岸又先出招——人工智能初创公司OpenAI公布新一代人工智能语言模型GPT-4。

与ChatGPT所用的模型相比,GPT-4不仅能够处理图像内容,而且在回答的细节和表述

方面更有价值和人情味。OpenAI称“它是迄今为止功能最强大的模型”;亦有评论认为它是硅谷在人工智能竞赛中“竞争下一代领导者”的重大赌注……全世界的人都在津津乐道它高达2.5万字的文字输入上限、强大的图像识别能力、更高的答题准确性、创造性文本的生成能力……

10%左右……

此外,GPT-4有一个质的飞跃——可以开始处理图像。不过,图像输入的权限尚未公开,所以目前还没看到更多例子来证明GPT-4在图像方面的处理能力。

有趣的是,GPT-4已经开始具备一点点的幽默感——它可以讲出一些模式化的、质量不佳的冷笑话。换句话说,它已经开始理解“幽默”这一人类特质。

### 令人期待但仍存缺陷

OpenAI表示GPT-4是其“最先进的人工智能系统”,并声称花费了6个月时间让它比上一代更安全。该公司通过改进监控框架、与敏感领域专家合作,来确保GPT-4所给出的答案的准确性和安全性。

与旧版本一样,GPT-4由深度学习技术开发而来,是在公开可用

和已被许可数据上进行的训练,并使用强化学习和人工反馈进行微调。不过,GPT-4是在全新的深度学习堆栈上训练的,并较上一代模型接受了更多数据训练和计算训练。

OpenAI还指出,GPT-4“离叛道”的可能性要小得多——对于不被允许的内容请求,GPT-4的响应可能性降低了82%。在OpenAI的内部对抗性真实性评估中,GPT-4得分比GPT-3.5高40%。

另外,当任务复杂性达到足够阈值时,两代模型之间的差异就会出现。整体来看,GPT-4比GPT-3.5更可靠、更有创意,并且能够处理更细微的指令。

在GPT-4推出后,其合作伙伴兼投资股东微软立即回应称“新的Bing正在GPT-4上运行,这是我们为搜索定制的”。事实上,随着

OpenAI对GPT-4及更高版本的更新,Bing也从中受益,Bing搜索引擎业已用上GPT-4,用户可以使用Bing编写文本、回复消息等。

不过,OpenAI并未公布有关具体模型大小和训练GPT-4的硬件详细信息。OpenAI创始人山姆·阿尔特曼表示,尽管GPT-4的未来非常令人期待,但是仍然存在缺陷,所生成的内容仍然存在一定的偏见性,也存在一定的“幻觉”,也就是GPT-4会创建“不正确但听起来似是而非”的虚构信息。

### 生成式AI时代到来

当元宇宙的酷炫残影还停留在人们的视网膜上,“ChatGPT”上知天文、下知地理的表现再次突破人类对AI能力上限的认知。但对大多数“吃瓜群众”来说,GPT热潮的背后,也藏着一份隐忧——在已

经到来的AIGC(人工智能生成内容)时代,我们要如何应对?

在中国科学院院士鄂维南看来,GPT的火热是OpenAI逐步验证、集中投入的结果。其中,“集中投入”这点非常重要。因为OpenAI的成功表明,以往那种“小作坊、项目制”的AI赋能方式正在成为历史,而“AI工程化、平台化”正成为释放人工智能技术红利的重要支撑。如果能够适应这种变化,AI在自然语言领域所取得的成功,有望在基础科学研究领域被复制。

“人工智能对于开发者来说是一种模式的变化,下一步对很多内容创作者来说,一样会面临工作模式的巨大变化。”商汤科技董事长兼CEO徐立指出,“人类正站在AI临界点上,其推动力就是生成式AI。生成式AI将来完全可以大幅降低软件开发门槛,将开发从‘特权’变成普惠行为。”

百度智能云主任架构师、百度AI中台总架构师谢永康分析,无论是ChatGPT还是之后产生出来的更多具备AGI(通用人工智能)能力的AI技术,更多应该是辅助人,与人找到共存的平衡点。

本报记者 郜阳

本报讯(记者 郜阳)昨天,多模态多任务通用大模型“书生(IntERN)2.5”发布,其图文跨模态开放任务处理能力可为自动驾驶、机器人等通用场景任务提供高效精准的感知和理解能力支持,向通用人工智能迈出坚实的一步。目前,“书生2.5”多模态通用大模型已在通用视觉开源平台OpenCVLab开源。

### 以文生图

“书生”由商汤科技、上海人工智能实验室、清华大学、香港中文大学、上海交通大学于2021年11月首次共同发布,并持续联合研发。“书生2.5”实现了通过文本来定义任务从而可以灵活定义不同场景的任务需求,并根据给定视觉图像和任务的提示性语句,给出相应指令或作答,进而具备通用场景下的高级感知和复杂问题处理能力,如图像描述、视觉问答、视觉推理和文字识别等。

研发人员介绍,在自动驾驶和居家机器人等通用场景下,“书生2.5”可辅助处理各种复杂任务。例如,在自动驾驶场景中可以大幅提

# 多模态多任务通用大模型“书生2.5”发布

## 具备通用场景下高级感知和复杂问题处理能力

升场景感知理解能力,准确辅助车辆判断交通信号灯状态、道路标志牌等信息,为车辆决策规划提供有效信息输入。

此外,“书生2.5”具备AIGC(人工智能生成内容)“以文生图”能力。可根据用户提出的文本创作需求,利用扩散模型生成算法,生成高质量、自然的写实图像。例如,通过生成各类真实的道路交通场景,如繁忙的城市街道、雨天拥挤车道、马路上奔跑的狗等,进而训练自动驾驶系统对场景的感知能力上限。

### 准确率高

记者了解到,“书生2.5”拥有30亿参数,是目前全球开源模型中ImageNet准确度最高、规模最大,同时也是物体检测标杆数据集COCO中唯一超过65.0mAP的模型。

“书生2.5”在图文跨模态领域

卓越的性能表现,来自于视觉、语音及多任务建模三大模型能力的有效融合,即InternImage-G通用视觉大模型、用于文本理解的超大语言预训练模型和用于多任务的兼容解码建模大模型。

值得一提的是,在视觉主流图像分类数据集ImageNet上,“书生2.5”仅基于公开数据便达到90.1%的准确率。这是除谷歌与微软之外,唯一准确率超过90.0%的模型——而谷歌与微软均未公开模型及额外数据集。该模型也在包括图像分类、物体检测、语义分割、图像描述、图文检索等20多个不同场景、不同任务的单模态和跨模态公开数据集里取得了最佳成绩。

当前,“书生”还在持续学习、不断进步,致力于推动多模态多任务通用模型技术的突破,驱动通用人工智能技术的创新应用生态。



“书生2.5”在自动驾驶场景中可以大幅提升场景感知理解能力

ICBC 中国工商银行 上海市分行

# 投资上海债 助力新未来

政府信用 / 安心回报 / 百元起购 / 灵活交易

期限:3年

代码:2305248

简称:23上海债01

销售期:3月21日-3月23日

认购渠道:募集资金用于市级的教育、文化、医疗卫生等民生类项目。

2023年上海市政府债券柜台发行