

OpenAI推出GPT-4o mini,号称“最具成本效益的小模型”

OpenAI推出GPT-4o mini,并声称这是“最具成本效益的小型模型”。

GPT-4o mini预计将通过让智能变得更加实惠,显著扩大使用AI构建的应用程序范围。

据OpenAI发布的消息,GPT-4o mini在MMLU(大规模多任务语言理解,衡量模型广泛的世界知识和解决问题的能力)上的得分为82%,比另外两款低成本竞品更高——Google的Gemini Flash得分为77.9%,Anthropic的Claude Haiku得分为73.8%。目前,GPT-4o mini在LMSYS排行榜上的聊天偏好方面优于GPT-4。

GPT-4o mini已经在几个关键基准上进行了评估,涉及文本和视觉的推理任务、数学推理和编码任务、多模态推理,结果均优于其他小型模型。

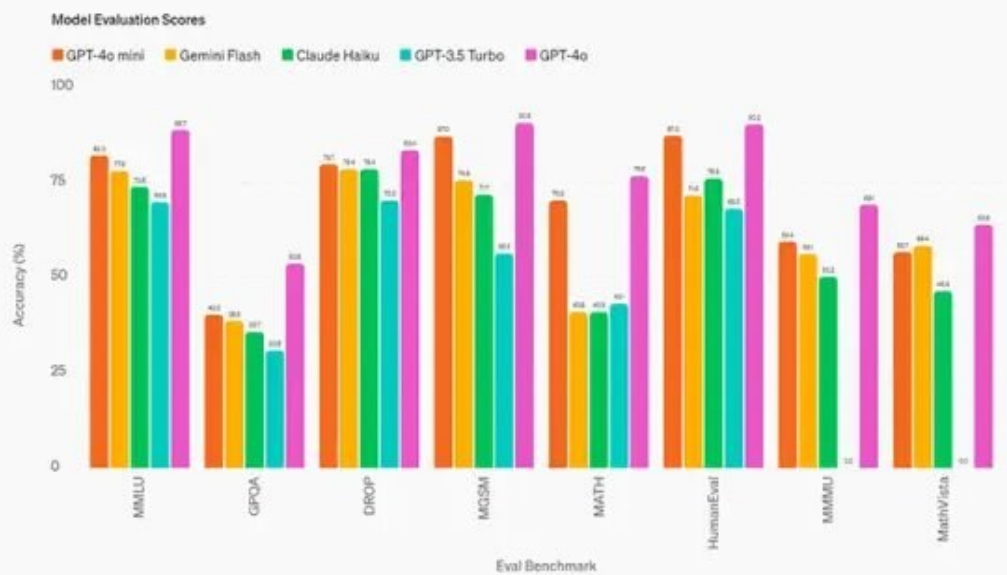
GPT-4o mini还在API中支持文本和视觉功能,未来还将支持文本、图像、视频和音频的输入和输出,并且拥有截至2023年10月的知识。得益于与GPT-4o共享的改进的tokenizer(编译器),GPT-4o mini处理非英语文本方面也更具成本效益。

GPT-4o mini成本更低、延迟更低,能够完成更多的任务,例如调用多个API(应用程序编程接口)、传递和记录大量完整的代码库或对话记录、更快速、实时地与使用者聊天等。

作为模型开发过程的一部分,OpenAI已经与一些值得信赖的合作伙伴合作,以更好地了解GPT-4o mini的使用场景和局限性。目前已与Ramp和超人等公司进行了合作,反馈发现在执行从收据文件中提取结构化数据或在提供线程历史记录时生成高质量电子邮件响应等任务时,GPT-4o mini的表现明显优于GPT-3.5 Turbo。

在此之前,考虑到价格因素,轻量化且廉价的模型“小而精”的AI模型可能更企业的受欢迎,小型模型的相关开发竞争激烈。

谈及为何没有更早地推出“小而精”的模型时,OpenAI的API产品负责人Olivier Godement说,“我们的使命是使最前沿技术、构建最强大、最有用的应用程序,我们当然希望继续做前沿模型,推动技术进步……但我们也希望拥有最好的小模型,我认为它会非常受欢迎。”



OpenAI称,开发过程中的每一步都考虑了安全性。开发时会过滤掉仇恨言论、色情信息、个人隐私信息等不宜模型学习和输出的内容,并采用自动和人工(超过70位来自社会心理学和错误信息等领域的外部专家)对模型进行了评估,以识别潜在风险。GPT-4o mini

是第一个应用指令层次结构的模型,有助于提高模型抵抗越狱、即时注入和系统即时提取的能力,这使得模型的响应更加可靠,并有助于使其在大规模应用中更安全地使用。

自2022年推出功能较弱的模型text-davinci-003以来,GPT-4o

mini的每token成本已下降99%。GPT-4o mini定价比之前的Frontier型号便宜一个数量级,比GPT-3.5 Turbo便宜60%以上。在ChatGPT中,免费版、Plus版和团队版用户已经能够使用GPT-4o mini替代GPT-3.5,企业用户也将从下周开始使用。

20美元! 苹果这项功能或收费! 分析师:意在将部分成本转嫁给用户

分析师称,苹果公司正酝酿一项重要举措,计划对其即将面世的高级人工智能系统Apple Intelligence向用户收取额外费用,预估价格将高达20美元。此举旨在进一步推动苹果利润丰厚的服务业务持续增长。

据悉,苹果计划在年内晚些时候正式在其部分设备上推出这款备受瞩目的人工智能系统。早在今年6月,苹果便宣布了这一人工智能系统的开发计划,并承诺将提升Siri语音助手的智能化水平,同时实现包括自动邮件撰写、

图像创作等在内的多项功能。

分析人士指出,虽然这些先进功能将分阶段逐步推出,且初期可能无法覆盖中国及欧洲等地区,但苹果有意对其中更高级的应用程序采取订阅收费模式。

Counterpoint Research的合伙人Neil Shah认为,人工智能技术的投资成本高昂,苹果此举意在将部分成本转嫁给用户。Shah在接受CNBC采访时透露,苹果或将Apple Intelligence的高级功能纳入现有的Apple One订阅套餐中,以每月10至20美元的价格提供给用户,以此吸引更多追求优质AI体验的用户。

继官宣“苹果智能”(Apple Intelligence)一个多月后,苹果这一人工智能(AI)系统终于落地终端设备。美东时间7月29日,该公司发布了苹果智能的首个iPhone AI版本。新软件目前只在iOS 18.1的开发者测试版中发布,仅面向支付了99美元/年的注册开发者。

随着iOS 18.1 Beta版的上线,注册开发者即日起就能体验到苹果AI的部分功能。

据悉,本次更新只支持M系和A17 Pro芯片,即在iPhone上仅适用于iPhone 15 Pro和iPhone 15 Pro Max,iPad和Mac也有类似的版本。除硬件要求外,开发者的地区设置需为美国,设备和Siri语言需为英语。

此次推出的苹果智能仅有部分功能上线,主要围绕文本生成、Siri和相册等模块。不过,此次更新尚未集成ChatGPT功能,苹果称该功能以及更多更新将在明年正式上线。

苹果基础大模型团队负责人Ruoming Pang强调,这些基础模型“并不是聊天机器人”,而是支持广泛的功能,包括摘要、写作帮助、工具使用和代码。

作为苹果AI的重要组成部分,文本生成功能的适用范围不局限于苹果官方应用,只要使用标准输入文本系统,在第三方应

用程序当中也能利用该功能进行文本总结、校对和重写。另外结合语音备忘录中已经上线的音频转录功能,文本生成系统还可以为录音生成摘要。

对于iOS 18.1最新测试版提供的AI功能,X网友很快发表了反馈。有网友称,“苹果智能帮我找到一种更好的方式来表达我真正想说的话。”在他的视频中,苹果智能将其对话中准备发送的内容修改得更加委婉。

此外,Siri的全面焕新也令人惊艳。有网友称赞苹果智能的新Siri界面,并称表示新版的Siri有“绝对的突破性”。

相册的更新则让用户可以用自然语言搜索特定照片,甚至是视频当中的具体时刻,并创造照片记忆。

7月29日,苹果还上线了一份苹果智能模型的技术报告,并表示在指令遵循、文本总结等任务上,苹果云端大模型取得了超过GPT-4的成绩。



曝苹果如期发布iPhone 16: 不等Apple Intelligence了

业内人士Mark Gurman透露,iPhone 16系列将会如期发布,这是一款没有预装AI功能的手机。

Mark Gurman表示,Apple Intelligence已在iOS 18.1测试版中上线,苹果需要很长时间完善AI体验,iPhone 16系列出厂不会搭载Apple Intelligence,9月份将如期发布。

苹果CEO库克曾表示,Apple Intelligence会在今年秋季上线。

据悉,今年iPhone 16系列将会推出标准版、Plus版、Pro版和Pro Max版。

配置方面,iPhone 16 Pro将配备6.3英寸显示屏,iPhone 16 Pro Max将配备6.9英

寸显示屏,iPhone 16标准版将配备6.1英寸屏幕,iPhone 16 Plus将拥有6.7英寸屏幕,后两款机型都是60Hz标准刷新率。

另外,4款机型都将支持Apple Intelligence,能深刻理解语言含义,支持邮件、备忘录、Safari浏览器、Pages文稿、Keynote讲演以及第三方App,大幅提升用户体验。

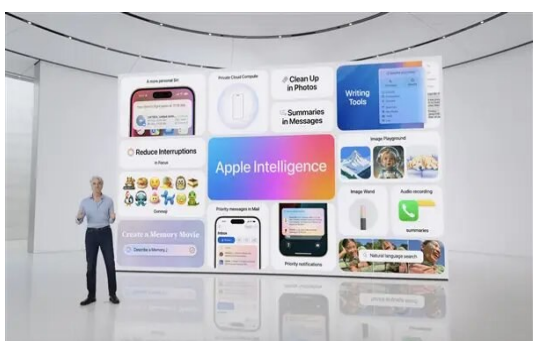
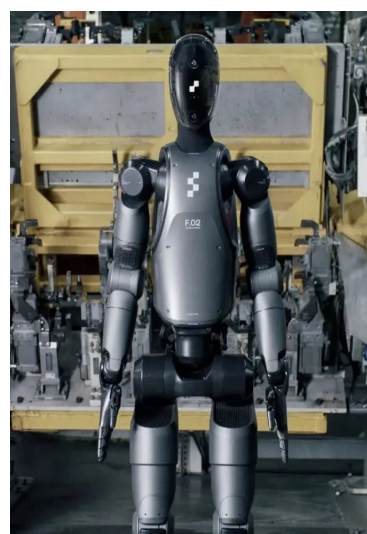


Figure AI第二代人形机器人已进入宝马工厂 称每天可工作20小时

近日,Figure AI推出了第二代人形机器人Figure 02,该机器人被誉为“世界上最先进的AI硬件”,



吸引了英伟达、OpenAI、微软和亚马逊创始人贝佐斯的投资关注,虽然AI智能尚未完全成熟,但人形机器人在工厂场景中的应用成为行业趋势,例如Figure AI与宝马合作,计划在汽车生产中部署人形机器人且已在宝马斯帕坦堡工厂的生产线上进行了测试,据介绍,该机器人可实现每天20小时以上的有效工作,有望进入工作场所和家庭。

Figure 02采用外骨骼结构,相比第一代Figure 01,内部布线更加集成,提高了结构刚度。外观上更接近人类,外露线缆减少,其GPU和CPU AI推理能力提升了3倍。机器人配备了与OpenAI合作训练的定制AI模型,支持语音交互,摄像头具备快速常识性视觉推理能力。

Figure 02通过板载麦克风和扬声器与人类进行对话,接入6个RGB摄像头的AI驱动视觉系统,并采用第四代手部设计,拥有16个自由度和与人类相当的力量。续航能力提升,内置2.25KWh电池组,能量高出第一代50%以上。

Figure AI此前因接入OpenAI大模型而受到关注,Figure 01能够通过语言描述所“看到”的画面,并进行快速响应。

虽然大模型与人形机器人的融合仍显初步,但大模型已开始用于任务规划,并在某些方面表现出良好效果,其他公司如优必选和开普勒也在推进人形机器人在工业场景中的应用,特斯拉CEO马斯克透露,Optimus人形机器人已在特斯拉工厂工作,预计明年数量将大幅增加。