



OpenAI 上新

长出“眼睛”“嘴巴”的 GPT-4o 发布

浙大人工智能专家:未来的大模型,将是一位超级“私人助手”

本报记者 朱高祥

美国当地时间5月13日22:00, OpenAI 举行了名为“春季更新”的线上发布会,宣布推出 GPT-4 的升级款模型 GPT-4o。

这场发布会堪称简陋,前后持续不到半个小时,也没有大屏 PPT,核心环节就是由首席技术官 Mira Murati 带着两位员工一起在现场展示新模型。

但这场发布会依旧惊艳,正如 GPT-4o 中的“o”(omni,意为全能),GPT-4o 长出了“眼睛”“嘴巴”,变得全知全能。

不少网友惊呼,电影照进现实,未来应用充满无限想象空间。

记者实测

支持输入图片和文件 更详细更全面

5月14日,GPT-4o 发布后,OpenAI 的首席执行官 Sam Altman 在社交媒体上发帖,仅有一个字“her”。

《her》是一部科幻电影的名字,影片中,人工智能系统“萨曼莎”不仅能够帮助男主西奥多完美地处理好工作,而且还是朵“解语花”。她拥有性感的声线,细腻的情感,并且风趣幽默,她能够和西奥多进行深度交流。

而“her”,也是很多人看到 GPT-4o 演示之后的最大感受,它似乎变得跟真人一样了。

昨天,为了验证 GPT-4o 的强大,本报记者特约在美国的同行,打开 ChatGPT 的 app 进行了实测。

通过交流,记者注意到,GPT-4o 与它的前任 GPT-3.5 最明显的不同在于界面,GPT-3.5 仅支持文字与语音输入,但 GPT-4o 可以输入文字、语音、图片以及文件等。

当记者询问 GPT-4o 可以做什么?它回答称,可以“回答问题”“语言翻译”“写作和编辑”“提供建议”“数学和编程”“数据分析”“创建图像”“实时信息查询”等。

对于“实时信息查询”,记者分别用“杭州今日天气”向 GPT-3.5、GPT-4 与 GPT-4o 进行了询问,只有 GPT-4 与 GPT-4o 可以回答。

本报记者拍摄了一张办公桌照片让 GPT-4 与 GPT-4o 进行描述,这两个大模型都可以在几秒钟之内给出解读。但对比来看,GPT-4o 总结得更为详细,也更为全面。

值得注意的是,与发布会上的演示不同,目前 ChatGPT 仅接入 GPT-4o 有限功能,尚不可以利用摄像头对现实场景分析,也不能在与其语音沟通时实时打断。

不过,OpenAI 已经宣布推出一款适用于 macOS 的桌面级应用,使用键盘快捷键就可向 ChatGPT 提问。

专家

印象最深的是 它能像“一个人”一样互动

在发布会上,OpenAI 现场展示了和 ChatGPT 的若干互动,包括:实时对话交互、语音多样化(应用户需求使用不同情绪、语调等)、视频指导做题、视频识别环境和人(包括人的情绪)、以桌面应用形式辅助编程、实时翻译等。

OpenAI 同时还放出了预先录制的展示样例若干,包括:两个 GPT-4o 交流和对唱、唱摇滚曲、在线会议应用、“毒舌”讽刺、视频识物并给出西班牙语单词、帮助面试准备、和狗互动等。

最让人惊讶的还是视频对话:用户打开摄像头,让 ChatGPT “看”到当下,并进行互动。

用前置镜头自拍,ChatGPT 可以识别用户的情绪,如“看起来很开心”;还能从用户背后的画面判断其身处环境,如“看起来你正在一个摄影棚中,背后有一些灯光,你的胸前还别着麦克风,可能在录制视频之类的”。

用后置镜头,ChatGPT 就可以和用户共享视角。如在语言学习的过程中,打开摄像头让 ChatGPT 用某种语言说出物品的名称。或者可以为视障用户提供指引,告诉用户“来了一辆计程车,现在招手吧”。

浙江大学人工智能研究所所长吴飞告诉钱报记者,GPT-4o 可以像“一个人”一样实时互动,是最让人印象深刻的地方。

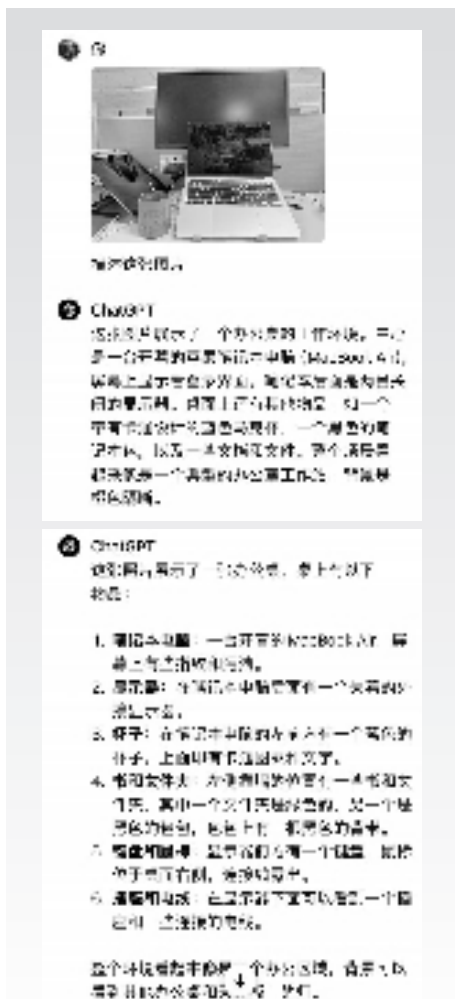
“自然语言交互也是 OpenAI 一直以来的理念,就是让人工智能像人一样能够交流。”吴飞说,在发布会上可以看到,与 GPT-4o 进行交流的时候,你不觉得对方是一个机器,就觉得对方是一个人,交互显得非常自然。

吴飞分析称,这些惊艳的产品表现,根本上源自于 GPT-4o 多模态大模型的技术进步,是跨文本、视觉和音频端到端地训练了一个新模型。

OpenAI 称这是其突破深度学习界限的最新举措。目前,GPT-4o 可以在短至 232 毫秒、平均 320 毫秒的时间内响应音频输入,与人类的响应时间相似。

吴飞认为,快速发展的多模态大模型给未来应用带来了丰富的想象空间。比如,也许不久的将来大模型不再仅仅是一个聊天机器人,而是将成为超级私人助手。不仅在工作上,在出游、订餐等生活的方方面面,大模型都可以快速地定制化生成用户所需的答案。

另外,GPT-4o 对周围环境的实时解读,为视障用户提供了更多方便,出行也将变得更为容易。



对同一张办公桌照片的解读,GPT-4o(下)比 GPT-4 更为详细全面