GPT-4o为OpenAI开启超级入口，对谷歌形成挑战？

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/27233.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

GPT-4o为OpenAI开启超级入口，对谷歌形成挑战？。 · 基于ChatGPT或GPT-4o，未来人类获得信息的方式很可能会改变，GPT-4o或为OpenAI开启了一个超级入口，这可能会对谷歌带来影响。接下来，OpenAI需要判断的是在产品上的极致体验是否是刚需。

GPT-4o在交互模式上是一个巨大进步。5月14日，蚂蚁集团副总裁、NextEvo负责人徐鹏对澎湃科技表示。2024年5月14日凌晨，
OpenAI向人们展示了其最新多模态大模型产品——GPT-4o，o代表omini，意为全能。

与现有模型相比，GPT-4o展现出了其在视觉和音频理解方面的出色技能。伴随着GPT-4o的到来，外界纷纷猜测，美国科幻电影《她》所描写的时代正一步步向我们靠近。2013年，电影《她》(《Her)》）中讲述了一个男人爱上了一个语音助手的故事。

与谷歌竞争原生多模态？

根据OpenAI首席技术官米拉·穆拉蒂（Mira Murati）的介绍，GPT-4o可以在音频、视觉和文本中进行实时推理，接受文本、音频和图像的任何组合作为输入，并生成文本、音频和图像的任何组合进行输出。它可以最短在232毫秒内响应音频输入，平均为320毫秒，这与人类在对话中的响应时间相似。

徐鹏在接受澎湃科技采访时表示，虽然OpenAI没有推出大众期待的GPT-5，但GPT-4o在交互模式上是一个巨大进步。和GPT-4相比，GPT-4o的最大区别在于，所有模态都集成在一个模型中，多模态整合更精细，延迟仅300毫秒左右，同时能够感知情绪、语气、表情，实现更自然的交互，这需要数据组织能力、聚焦突破能力、工程优化能力，也扩大了人们对于交互的想象空间。

徐鹏认为，自去年12月谷歌推出原生多模态Gemini模型之后，OpenAI就在为原生多模态领域的竞争做准备了。他所在的蚂蚁集团在今年年初判断原生多模态技术方向后坚定投入，目前也正在研发全模态数字人和全模态智能体的产品。

所谓原生多模态，即从一开始就使用多种模态（例如音频、视频和图像）训练模型，而不是拼凑多模态模型。

徐鹏表示，OpenAI的目标是实现多模态深度结合，早在GPT-3时代，其推出的自动语音识别系统Whisper就是一个前期研究。把语音、图像、视频、文字等各种模态的数据放在统一表征框架下，对于实现他们眼中的API（应用程序编程接口）是一个非常自然的方式，因为人也是多模态理

解和交互的智能体。

猎豹移动董事长兼CEO傅盛表示，虽然GPT-4o让人工智能从业者比较失望，但他也指出，GPT-4o相当于把一系列引擎结合在一起，比如图片、文字、声音，这样用户就不需要来回切换了。最重要的是这次发布的语音助手，由于使用了端到端的大模型技术，它能够实时感知情感变化，在该插话的时候插话，其实这才是大模型的未来。

GPT-5可能还要难产一段时间？

徐鹏介绍，原生多模态有三个特点：一是端到端做训练，二是实现了多模态的统一训练，能读、能听、能说，三是实现复杂推理。把图像、文字、语音、视频编码编到一个模型里，在模型里它们有统一的表征，这些数据一起送给模型训练，模型就会学到各个模态，只要它们的信息是相关的，内部的表征实际上是非常接近的，这样在生成时也会比较灵活。徐鹏表示，内部表征已经融合，所以GPT-4o可以最快的速度输出生成的语音，实现低延时的丝滑交互。OpenAI的工程能力确实令人赞叹，模态这么多，输入Token数非常多，还能够以两三百毫秒的延迟输出，这在工程上是难得的进展。

对于此次GPT-4o模型，目前，OpenAI高管并未透露训练GPT-4o模型时使用了什么样的数据，也未透露OpenAI能否利用较少的算力训练该模型。

科技投资人、华创资本创始合伙人熊伟铭告诉澎湃科技，尽管OpenAI在此次发布会上并未透露太多关于GPT-4o模型训练的技术细节，但可以猜测出，这类端到端的大模型技术的实现，背后依赖了强大的算力支持。肯定是大力出奇迹，这一点美国的算力市场确实成熟很多，资本市场也支持大规模的算力投入。熊伟铭说。

傅盛认为，如果不计成本地叠加参数，提高所谓的大模型能力，这条路肯定会遇到困难。他预计，GPT-5可能还要难产一段时间。

超级入口已经开启？

OpenAI官网显示，目前，GPT-4o的文本和图像功能开始在ChatGPT中免费推出，Plus用户可以享受到5倍的调用额度。新版语音模式将在未来几周向Plus用户推出，同时也将会在API（应用接口）中向小范围推出对GPT-4o的新音频和视频功能的支持。

在熊伟铭看来，OpenAI的产品策略一方面能够吸引免费的用户使用，可以收集用户大量的数据投喂给模型训练，有助于进一步改进产品，这样交互的大模型用户数据将会非常丰富。另一方面，能够培养用户的付费意愿，也是商业化的一种尝试。

我觉得OpenAI的这种尝试，可能会改变国内一些用户使用软件的习惯。大家可能会愿意付费在AI平台的使用上。熊伟铭说。

徐鹏认为，OpenAI免费为用户开放服务看中的是基于GPT-4o原生多模态的能力，未来更多企业可以在GPT-4o的基础上开发出更自然的垂类交互产品。

过去一周，外媒不断有消息传出OpenAI将推出AI搜索产品，虽然OpenAI没有推出搜索引擎，但是徐鹏认为基于ChatGPT或GPT-4o，未来人类获得信息的方式很可能会改变，GPT-4o或为Open

AI开启了一个超级入口，这可能会对谷歌带来影响。接下来，OpenAI需要判断的是在产品上的极致体验是否是刚需。

信也科技副总裁、大数据及AI负责人陈磊告诉澎湃科技，从技术上看，GPT-4o的发布具有跨时代意义，真正实现了多模态交互，更需要关注的是后续商业化中究竟如何持续落地。语音识别、语音生成不是最难的，最难的是推理归纳。GPT-4o做题比以前做题难度更高了。把算法调到一定程度是能实现交互流畅性的，但像人一样思考、推理、归纳、总结，这是更高智能的体现。

陈磊同时表示，国内还在对标GPT-4时，OpenAI推出了GPT-4o。行业需要思考的是在持续追赶中如何差异化竞争。我们始终在追，追到一定程度发现推出了新一代产品，我们始终感觉在落后，所以要调整心态，寻找另外一条路。陈磊说。

傅盛认为，OpenAI发布GPT-4o模型应用，正说明了大模型在应用层面大有可为，大模型的能力会不断迭代，但最终能把大模型用好的，还是应用。

作者：张静,喻琰,方卉 来源：澎湃新闻

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发