

## OpenAI 宣布 ChatGPT 大幅降价 对于竞争对手来说是个坏消息

北京时间 3 月 10 日早间消息，为当前人工智能热潮提供支撑的最重要“原材料”的价格正在快速下降，这将有助于这项技术更快地进入主流。然而，这也威胁到了那些希望从这股热潮中获利的创业公司的财务状况，并可能导致行业主导权集中在一小部分公司的手中。

这里的原材料指的是大语言模型（LLM）的处理能力。这些模型支撑着 ChatGPT 和微软新必应搜索等服务。



运行这些模型原本需要高昂的算力成本，因此可能会严重拖累模型的广泛应用。搜索引擎 You.com

首席执行官、知名计算机科学家理查德 索切 (Richard Socher) 表示，就在几周前，You.com 使用人工智能提供搜索服务的成本还比传统互联网搜索高出 50%。但到上月底，由于大语言模型公司 OpenAI、Anthropic 和 Cohere 之间的激烈竞争，这一成本差异下降到只有约 5%。

几天后，OpenAI 发布了一项新服务，让开发人员可以直接使用 ChatGPT，并将使用这项技术的价格削减了 90%。

这对客户来说是好事，但对 OpenAI 的竞争对手来说可能是毁灭性的。包括 Anthropic 和 Inflection 在内的许多公司已经完成或正在融资，以支撑各自大语言模型的发展。

很少有像大语言模型这样的技术能如此迅速地从实验室研究直接进化到大规模商用，这也促使研究者纷纷将实验室环境的开发过程“工业化”。性能上的大部分提升以及成本的下降主要来自运行大语言模型的底层计算平台的优化，以及对模型训练和运行方式的改进。

从某种角度来看，硬件成本的大幅下降有利于所有的市场参与者。这其中包括为满足最新人工智能模型要求而设计的高性能芯片，例如英伟达 H100 GPU。微软在其 Azure 云计算平台上运行 OpenAI 的模型，并向其他大语言模型公司提供相同的高性价比硬件支持。

然而，大语言模型既是科学，也是艺术。OpenAI 表示，自去年 12 月以来，ChatGPT 对于处理查询的方式进行了“一系列系统范围的优化”，使成本降低了 90%，最终带来了面向用户的大幅降价。

训练大语言模型需要花费数千万美元，而处理这类任务的技术也在快速变化。至少在短期内，一小部分具备模型开发和训练经验的人才将可以获得更大的优势。

当最优秀的技术被广泛理解和采用时，早期参与者可能已经获得了先发优势。微软云计算和人工智能部门负责人斯科特 格里 (Scott

Guthrie) 提到了 GitHub Copilot 等新服务。该服务于去年夏季推出，向软件开发者提供代码建议。在被广泛使用后，这类服务将会迅速优化。他本周在摩根士丹利的一场投资者会议上表示，来自这类服务用户的“信号”很快就将成为产品重要的差异化点。

OpenAI 竞争对手的主要希望在于，提供额外的服务，帮助开发者和大企业客户更方便地使用大语言模型，以及针对细分市场探索满足特定业务需求的模型。

例如，以色列创业公司 AI21 Labs 本周发布了最新的大语言模型，同时也发布了一系列 API（应用程序接口），以提供文字概要或改写等更高级的服务。

AI21 联席首席执行官奥利 格申（Ori Goshen）表示，大部分公司不会使用类似 ChatGPT 的通用模型，而是需要针对金融或医疗等行业训练的模型，或是基于某个公司自有数据训练的模型。

他认为，大语言模型目前还处于初级阶段，还有很多工作需要完成，例如减少模型说假话的倾向，以及防止模型产生“幻觉”，提供与事实无关、似是而非的答案。如果希望取得成功，人工智能公司还需要持续进行前沿探索。

不过目前的事实是，这些生成式人工智能的基础成本正在大幅下降。OpenAI 的降价是一个信号，表明这项技术将以非常快的速度进入大规模商用。但这同时也带来了警示，表明这个行业未来可能不会有太多公司有实力参与。

本文链接：<https://dqcm.net/zixun/167841381911810.html>