ChatGPT mania may be cooling, but a serious new

industry is taking shape

Three forces will shape the business of generative AI



THE FIRST wave of excitement about generative artificial intelligence (AI) was like nothing else the world had seen. Within two months of its launch in November 2022, ChatGPT had racked up 100m users. Internet searches for “artificial intelligence” surged; more than $40bn in venture capital flowed into AI firms in the first half of this year alone.

The craze for consumer experimentation has since cooled a little: ChatGPT use has fallen and fewer people are Googling “AI”. Son Masayoshi, a Japanese investor notorious for diving into already frothy markets, is thought to be interested in investing in OpenAI, ChatGPT’s creator. But a second, more serious phase is beginning. An entirely new industry centred on supercharged AI models is taking shape. Three forces will determine what it eventually looks like—and whether OpenAI stays dominant, or other players prevail.

生成式人工智能（AI）的带来的第一波热潮是前所未有的。ChatGPT于2022年11月推出后的两个月内就吸引了1亿用户。人工智能”的互联网搜索量激增；仅今年上半年就有超过400亿美元的风险投资流入AI公司。

此后，关于消费者实验的热潮稍微有所冷却：ChatGPT的使用量有所下降，搜索“AI”的人数也减少了。日因涉足已经存在泡沫的市场而臭名昭著的日本投资者孙正义，被认为有兴趣投资ChatGPT的创造者OpenAI。但是，第二个更为严重的阶段开始了。一个以增强AI模型为中心的全新产业逐渐形成。三股力量将决定其最终的形态——以及OpenAI是否保持主导地位，或者其他参与者占上风。

The first factor is computing power, the cost of which is forcing modelbuilders to become more efficient. Faced with the eye-watering costs of training and running more powerful models, for instance, OpenAI is not yet training its next big model, GPT-5, but GPT-4.5 instead, a more efficient version of its current leading product. That could give deep-pocketed rivals such as Google a chance to catch up. Gemini, the tech giant’s soon-to-bereleased cutting-edge model, is thought to be more powerful than OpenAI’s current version.

High computing costs have also encouraged the proliferation of much smaller models, which are trained on specific data to do specific things.Replit, a startup, has trained a model on computer code to help developers write programs, for instance. Open-source models are also making it easier for people and companies to plunge into the world of generative AI.According to a count maintained by Hugging Face, an AI firm, roughly 1,500 versions of such fine-tuned models exist.

第一个因素是计算能力，其成本花费迫使模型构建者变得更加高效。例如，面对训练和运行更加强大的模型所需要的让人难以承受的成本，OpenAI尚未训练其下一个模型-CPT 5,而是训练一个基于当前领先产品更加高效的模型 - GPT-4.5.这可能给财力雄厚的竞争对手比如Google 迎头赶上的机会。Gemini是这家科技巨头即将发布的尖端模型，被认为比Open AI当前版本刚加强大

高额的计算成本另一方面促进了小型模型的激增，这些模型利用特定的数据进行训练来完成特定的事情。例如，一家初创公司 Replit, 已经利用计算机代码训练出了一个模型来帮助开发者编写大吗。开源模型也让人们和公司投入到人工智能的世界变得更加简单。根据人工智能公司Hugging Face 的统计，此类微调模型的版本大约有 1,500 个版本。

All these models are now scrambling for data—the second force shaping the generative-AI industry. The biggest, such as OpenAI’s and Google’s, are gluttonous: they are trained on more than 1trn words, the equivalent of over 250 English-language Wikipedias. As they grow bigger they will get hungrier. But the internet is close to being exhausted. Many model-makers are therefore signing deals with news and photography agencies. Others are racing to create “synthetic” training data using algorithms; still others are trying to work with new forms of data, such as video. The prize is a model that beats the rivals.

Generative AI’s hunger for data and power makes a third ingredient more important still: money. Many model-makers are already turning away from ChatGPT-style bots for the general public, and looking instead to fee-paying businesses. OpenAI, which started life in 2015 as a non-profit venture, has been especially energetic in this regard. It has not just licensed its models to Microsoft, but is setting up bespoke tools for companies including Morgan Stanley and Salesforce. Abu Dhabi plans to establish a company to help commercialise applications of Falcon, its open-source AI model.

所有现在这些模型都在争抢数据-这就是第二股想要重新塑造生成式人工智能行业的力量。科技巨头，比如Google和OpenAI,非常贪婪，他们基于超过了1trn 个单词进行训练，这相当于超过250个的英语维基百科。随着他们变得越来越强大，他们也变得越来饥饿。但是互联网已经发展已经到瓶颈期了。因此，许多模型制作者正在与新闻和摄影机构签署协议。其他人正在竞相使用算法创建“合成”训练数据；还有一些人正在尝试使用新形式的数据，例如视频。奖品是击败竞争对手的模型。

生成式人工智能对数据和潜力的渴望使得第三个要素要素变得更加重要：金钱。许多模型制作者已经不再为公众提供 ChatGPT 式的机器人，而是转向付费企业。OpenAI 于 2015 年作为一家非营利性企业成立，在这方面尤其充满活力。它不仅将其模型授权给微软，还为摩根士丹利和 Salesforce 等公司建立定制工具。阿布扎比计划成立一家公司，帮助其开源人工智能模型 Falcon 的应用商业化

Another approach is to appeal to software developers, in the hope of getting them addicted to your model and creating the network effects that are so prized in tech. OpenAI is offering tools to help developers build products using its models; Meta hopes that LLaMA, its open-source model, will help create a loyal community of programmers.

Who will emerge victorious? Firms like OpenAI, with its vast number of users, and Google, with its deep pockets, have a clear early advantage. But for as long as computing power and data remain constraints, the rewards for clever ways around them will be large. A model-builder with the most efficient approach, the most ingenious method to synthesise data or the most appealing pitch to customers could yet steal the lead. The hype may have cooled. But the drama is just beginning.

另一种方法是吸引软件开发人员，希望让他们沉迷于你的模型并创造技术非常珍贵的网络效应。 OpenAI 正在提供工具来帮助开发人员使用其模型构建产品； Meta 希望其开源模型 LLaMA 能够帮助创建一个忠诚的程序员社区。

谁将取得胜利？像 OpenAI 这样拥有大量用户的公司和谷歌这样财力雄厚的公司拥有明显的早期优势。但只要计算能力和数据仍然受到限制。