

1万亿美元！ 全球市值最高芯片公司诞生了

5月30日美股盘前，英伟达市值冲破万亿美元，成为美股有史以来首个市值触及1万亿美元的芯片公司，创造了历史。

每日经济新闻报道，在最近的人工智能浪潮中，英伟达公司无疑是站在风口浪尖的一家公司，其风头之劲，甚至在最近压过了ChatGPT的东家微软。而英伟达之所以能独领风骚，一个关键原因在于其广受人工智能领域追捧的芯片产品，即A100芯片及更高一代的H100芯片，目前这些高端芯片及相应的显卡已是一卡难求。

微软亚洲研究员的高级研究员张弋近期在播客节目中感叹，现在居然到了整个地球都拿不出足够的A100芯片的奇怪场景。一年前，几乎无人料到这一情形。英伟达2020年推出的A100芯片现在是有价无市，而乘着ChatGPT爆红的H100更是被大公司疯狂抢购。这也让英伟达的业绩一路高歌，股价更是节节攀升。

人工智能领域的初创公司Core Weave创始人兼首席执行官Brannin McBee不禁感叹：H100是地球上最稀缺的工程资源之一。这句表态足以让人窥得英伟达现在的盛景。

英伟达成全球市值最高芯片公司接近于两个台积电

5月30日晚，美股开盘，英伟达股价上涨3.8%至404.25美元/股，截至发稿，英伟达报410.06美元，涨5.29%，其市值已突破1万亿美元，达到1.014万亿美元。

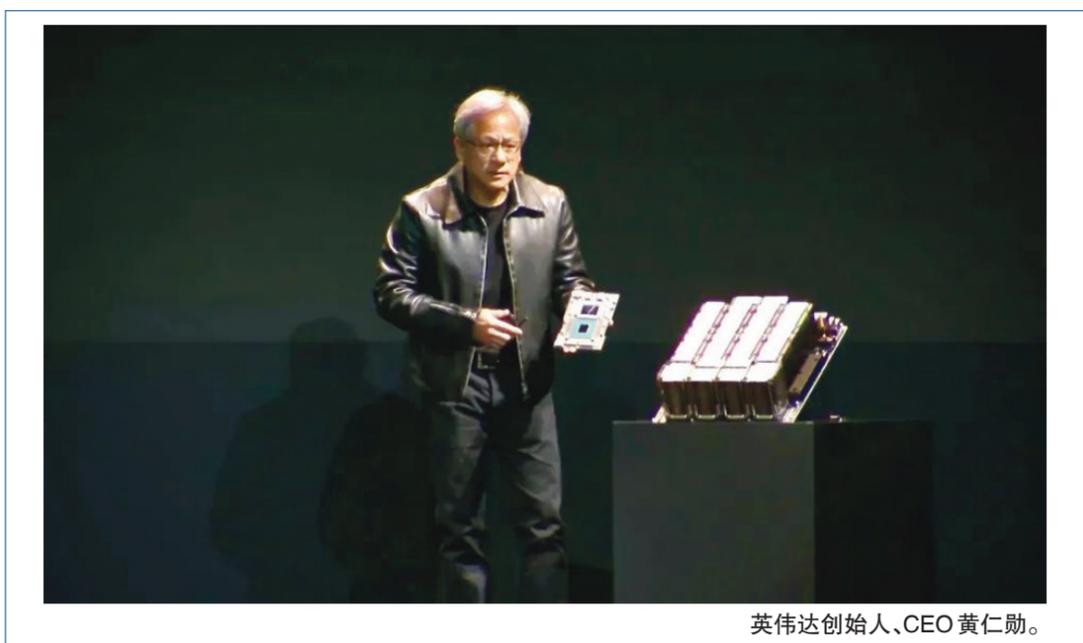
根据companiesmarketcap网站显示，英伟达1万亿美元市值排名全球第6，也是目前市值最高的芯片公司，接近于两个台积电（5340亿美元），与亚马逊的市值仅相差约两千多亿美元，今年以来英伟达股价涨幅约180%。

消息面上，据外媒报道，近期投资银行摩根大通在其投资报告中称，凭借GPU和网络产品等硬件产品，英伟达今年将在人工智能（AI）产品市场中占据高达60%的份额。

这一趋势也反映在了英伟达一季度财报当中，5月25日，英伟达公布今年一季度业绩，收入71.9亿美元，同比下降13%；净利润20.4亿美元，同比上涨26%。但同期，其数据中心业务营收达到创纪录的42.8亿美元，同比增长14%，这部分营收占其总营收的60%；游戏业务营收为22.4亿美元，同比下降38%，占其总营收的31%。

在上周发布了显著好于业内预期的收入预测后，英伟达股价一夜“狂飙”近30%。

业绩电话会上，英伟达向分析师表示，众多云公司竞相部署AI芯片，个人电脑（PC）的图形处理单元（GPU）终端需求一季度“稳



英伟达创始人、CEO 黄仁勋。

固”。业绩增长正来自数据中心业务。公司锁定了数据中心芯片的大幅增长，计划下半年大幅增加供应。

同时，该公司对下一季度的业绩做出指引，预测二季度的销售额将达到110亿美元，上下浮动2%，远高于外界预期的72亿美元。这也导致英伟达股价在盘后交易中飙升25%，增加了近2000亿美元的市值。

据财联社，分析师指出，对英伟达而言，这波AI热潮带来的商机将远比加密货币来得重要和持久。英伟达的芯片和软件可以满足生成式AI的计算密集需求，其产品的丰富性在业内无人能敌。据瑞银分析师估计，开发聊天机器人ChatGPT需要用到大约1万枚英伟达GPU。

今年人工智能的火热，带来芯片需求的大爆发。作为一站式解决方案的提供商，英伟达提供可以帮助AI产品训练大量文本、图像和视频的GPU，在AI大模型竞赛中掌握着供应算力的“命门”，由此成为今年AI概念热潮中的一大赢家。

目前，英伟达是GPU市场的主导者，全球独立显卡市占率高达80%，其高端GPU如H100、A100和V100等，占据AI训练市场绝大部分的份额。由于GPU能够为大型语言模型提供算力，市场对于其需求呈上升趋势，此前市场上曾有英伟达芯片供应短缺的消息传出。

推出AI超级计算机：内存扩大500倍！

芯片制造商今年的股价上涨很大程度上受到人工智能对算力需求的推动。英伟达正站在“超级算力”的风口上。

据澎湃新闻，5月29日，英伟达CEO黄仁勋在周一举行的

COMPUTEX大会开幕演讲中，发表了超级计算机DGX GH200，黄仁勋称它“集成了英伟达最先进的加速计算和网络技术。”

DGX GH200是一款大内存AI超级计算机，可用于支持生成式AI、数据处理的巨型模型。同时它也是第一款将Grace Hopper超级芯片与NVLink Switch System配对使用的超级计算机，此种互连方式能够使其系统中的所有GPU作为一个整体协同运行。

据介绍，DGX GH200共链接256个Grace Hopper超级芯片，能够提供1exaflop的性能、具有144TB的共享内存，相较2020年推出的上一代DGX A100，内存扩大了近500倍。

合作伙伴方面，谷歌云、Meta和微软有望首批接入DGX GH200，将其用于生成式AI产品。下一步，英伟达还打算将DGX GH200设计作为蓝图，提供给云计算公司和其他超大规模企业。

在此基础上，英伟达还计划打造基于DGX GH200的AI超级计算机“Nvidia Helios”，这款超级计算机将配备四个DGX GH200系统，包含1024个Grace Hopper超级芯片，预计将于今年年底上线。

据第一财经，英伟达还在构建一台超大规模生成式AI超级计算机Israel-1，将被部署在英伟达以色列数据中心。

黄仁勋表示，人们现在正处于一个新的计算时代的转折点，加速计算和人工智能已经被世界上几乎所有的计算平台和云服务厂商所接受。

“生成式AI是我们这个时代最重要的计算平台。有了生成式AI，现在每一个人都可以成为程序员。”黄仁勋表示，“他们只需要对

着计算机说话，就会有新的功能和应用层出不穷地涌现。”

英伟达凭什么赌赢了大趋势？
世上芯片千千万，为何只有英伟达的芯片成为人工智能领域独一无二的玩家？而英伟达这一向来在显卡称霸的公司，又为何能在深度学习和人工智能领域打下这么大的一片江山？

据财联社，1999年，初露头角的英伟达首次推出了GPU这一概念。在此之前，包括英特尔在内的CPU厂商都坚信图形处理是CPU的活，CPU干的事情越多越好，将图形工作独立到另一附属处理器上的想法十分鸡肋。

当时，图形应用领域又以搞游戏的日本厂商话语权最重。日系主机的CPU很强，大部分开发工作都集中于CPU之上，因此GPU并没有得到多少市场空间。

转机在于，不服气的微软想要冲击属于日本厂商的行业领先地位，其开发出了Direct X这一标准化的API图形接口，此后大量的图形功能从CPU里面移植出来，转到GPU之上。加上微软另一产品Xbox的推出，其CPU、GPU各司其职的搭配，打破了行业内CPU芯片一家独大的局面。

而英伟达是当年硬件领域唯一跟着微软旗帜前进的公司，并在GPU这条道上一路走到黑。

此后，微软又推动了另一次变革，引入统一渲染技术，即让GPU将图形绘制的顶点计算和之后的渲染两个步骤进行合并。它和显卡领域另一知名公司ATI合作了GPU Xenos，成功应用了这一技术。

统一渲染只是图形应用上的一步，但却给英伟达带来了完全不同的发展路径，可以说是英伟达在

后来GPU发展，甚至介入深度学习领域中的起点。

在看到统一渲染架构之后，英伟达果断的把自己从前的GPU架构推倒重来。其GPU流处理器被进行了细致的分组，变成一个个小型流处理器且能单独运行，解决了流处理器此前被绑定无法独立运行而被迫闲置的问题。

这奠定了英伟达后来革命性CUDA架构的出世。由于英伟达的流处理器是很独立且标准的单元，极易控制和调度，这让原本只能串行处理的任务可以被并行处理。这让编程难度大大降低。

与此同时，英伟达的竞争对手ATI却因为早期没有投入硬件架构变革，因为沿用过去的串行设计，沉没成本越来越高，让其革新变得越来越难且越来越贵，最后成功被英伟达挤出显卡市场。

此后，英伟达又在2017年引入了Tensor Core计算单元概念，其专门为深度学习而设计，支持更低精度的运算从而大幅节省了模型算力。

这一专用的加速单元很大程度上排挤了CUDA处理深度学习的空间，但也同时打了英伟达竞争对手一个措手不及，让AI专用芯片也不再吸引人。于是，英伟达GPU机缘巧合地成为AI领域最被认可的硬件。

2003年，“快速迭代，不断试错”的英伟达搞了一个不受欢迎的项目。其开发了一款Soc芯片，将基于ARM架构的CPU与自己的GPU集成在一起。自Soc芯片之后，英伟达每隔几年都发布一些芯片。2015年，其推出了Tegra K1，使用Arm公版CPU和自己的开普勒架构GPU，但由于功耗和发热皆不尽人意，对大部分使用者来说很是折磨。

但业内人士却对这些挫折十分认可。一位投资人曾指出，英伟达在守住GPU基本盘的同时，不断在新的领域伸出触角，并让无数买它显卡的人陪它分摊成本。他还称赞道，虽然英伟达的很多东西，如CUDA在一段时间内看不到落地的场景，但在试错过程中它建立起了完整的生态，并在一股新风向袭来时，成功站上了风口。

这也是英伟达GPU打败其他芯片，成功吃下AI红利的一个原因。一方面，GPU的通用性更佳，比专用芯片更适应变化；另一方面，英伟达拥有完整生态，让它的GPU成为当下最合适的那个选择。

事实上，当AI一瞬间爆发，行业内的企业无奈发现，GPU是简单高效运行生成式AI模型的最好选择，一个本用来玩游戏的GPU不太可能切换去跑AI程序，目前只有英伟达的GPU能够做到运行AI模型。

加入AI竞赛 TikTok测试聊天机器人Tako

据媒体报道，社交媒体应用TikTok正在测试人工智能（AI）聊天机器人Tako。TikTok称，Tako可与用户就短视频进行交流，并帮助他们发掘内容。

综合路透社、彭博社报道，TikTok于25日表示，正在菲律宾的选定用户群中测试这款名为“Tako”的AI聊天机器人，但仍处于初步探索阶段。TikTok的发言人说：“在选定的市场，我们正在推动TikTok上的搜索和发掘（内容）的新方法。我们期待从我们的用户社区中学习，同时继续创造一个安全的地方，带来娱乐、激发创造力并推动文化。”

总部位于以色列的AI科技公

司Watchful Technologies表示，已经在苹果移动设备上某些版本的TikTok应用程序中发现了这款被称为Tako的AI聊天机器人。

Watchful Technologies的研究员布查克（Daniel Buchuk）说，与定位为多用途的聊天机器人ChatGPT不同，Tako感觉更像是一个导航助手，重点是鼓励用户观看更多视频。

根据相关截图和视频信息，Tako在TikTok程序中显示为一块幽灵形状的图标，用户可以在观看视频时点击它，进行文本对话，并在Tako的帮助下寻找内容。

据了解，TikTok还设置了一个免责声明，称Tako是一个实验性的聊

天机器人，答复可能不准确；TikTok将出于安全目的审查与Tako的对话，并警告用户不要与Tako分享私人信息。

有媒体4月曾报道称，TikTok正在测试一种生成式AI工具，允许用户创建头像（avatar）。此外，上海《科创板日报》也曾在2月的报道中称，TikTok中国母公司字节跳动的人工智能实验室（ByteDance AI Lab）有开展类似ChatGPT和AIGC（生成式AI）的相关研发，未来或为旗下XR（扩展现实）平台PICO提供技术支持。

据路透社，自去年微软公司推出聊天机器人ChatGPT后，触发了各家科技公司的AI研发竞赛，包括



TikTok的竞争对手Snap在内。Snap于今年2月早些时候推出了“My AI”功能，且是使用OpenAI的GPT技术开发的。

报道还称，上个月提交给美国专利和商标局的信息显示，TikTok

已经提交了“TikTok Tako”的商标申请，所属归类包括“用于人工制作人类语音和文字的计算机软件”。