作者：万晨

编辑：郑璇

来源：极客公园

图片来源：由 Maze AI 生成

近日，英伟达公布了2023财年及其第四季度的财报。在加密货币低迷、消费需求疲软和去库存化的压力下，英伟达&#039；美国2023财年的总收入与上一财年基本持平，约为270亿美元。在…之中第四季度营收为60.5亿美元，比去年同期下降21%。

尽管如此，其业绩仍优于分析师&#039；之前的预期。财报发布当天，英伟达&#039；美国股价上涨14%，市值达到5800亿美元。实际上人&#039；美国对英伟达的乐观情绪已经蔓延了几个月。自今年1月以来，英伟达的市值涨幅高达60%。

感谢OpenAI。其发布的ChatGPT、DALL-E2等大语言模型将生成性AI引入大众&#039；ssight.——几乎所有的软件都会被AI重塑。黄仁勋甚至将其比作&quot人工智能的iPhone时刻&quot。

至此，时代的风口突然从元宇宙和web3切换到了生成式AI，FAAMG等硅谷巨头争相准备随时应战。和英伟达，牢牢成为&quot最大的军火商&quot在这个时代的战争中。

作为&quot人工智能超周期心脏跳动&quot，英伟达&#039；sGPU(图形处理芯片)是训练和操作机器学习模型的最佳选择。因此，它被视为&quot2023年云资本支出重心向人工智能转移的最大受益者&quot。

其实这已经不是英伟达第一次拿时代的风车3354来加速计算、深度学习、挖掘、元宇宙了。英伟达曾多次踩在中年的风口浪尖上。。成立短短30年，芯片世界风云变幻，当年和90家显卡厂商一起败下阵来的创业公司，早已成为市值最高的芯片霸主。

英伟达屡次卧薪尝胆。没有舵手黄仁勋的战略眼光，——总能准确预测下一次技术变革，提前下手。在最近的财报电话会议上，黄仁勋透露：这一次，他提前看到了未来及其相应的战略布局。面对大语言模型加持的生成人工智能一个&quot核弹工厂&quot远远不能提供&quot武器&quot。

01ChatGPT大战背后的战争之主

从去年11月底开始，OpenAI让人们看到了&quot普通情报&quot。依靠大语言模型的ChatGPT，虽然ChatGPT本身没有知识和智慧，但却表现出惊人的思维链和各种能力的自发涌现。但是它已经达到了&quot让你觉得它有知识甚至智慧&quot。不久前，在加州大学伯克利分校哈斯商学院(HaasSchoolofBusiness)的炉边聊天上，黄仁勋兴奋地评论道，ChatGPT将开启科技行业的一个新时代。也是人工智能和计算行业历史上最奇妙的事情。他说，&quot你最后一次看到这样一种多功能的技术是什么时候，它可以解决问题，并经常在许多方面给人带来惊喜？它可以写一首诗，填写电子表格。，您可以编写SQL查询并执行它们，您可以编写Python代码.对于很多一直致力于此的人来说，我们一直在等待这一刻，这就是人工智能的iPhone时刻。。我现在可以将它作为一个API，并将其连接到电子表格、PPT和各种应用程序，这有可能使一切变得更好。"

这是「AI 将重塑所有软件」的际遇生成式AI要想像ChatGPT一样展现各种通用才能，就必须依赖GPT3.5这样的底层大语言模型，人们把它比作移动互联网时代的Android或者iOS。因此大语言模式已经成为大厂和创业公司的战场。

无论是&quot建筑&quot这么大的模型或者运行这么大的模型，需要很大的计算能力和上千个GPU。据报道OpenAI用了1万个NVIDIAGPUs来训练ChatGPT。花旗集团估计，ChatGPT的使用可能会在12个月内为英伟达带来30亿至110亿美元的销售额。

之前《中国电子报》采访业内人士说，&quot大模型技术涉及AI开发、推理、训练的方方面面。所谓的&quot大&quot模型主要是因为参数和计算量大，需要更多的数据和更高的计算支持。对于GPU厂商来说，大型号是值得期待的计算红利，尤其是多功能的NVIDIA。"

在全球范围内，计算芯片领域主要有两大玩家，NVIDIA和AMD。从市场份额来看，英伟达远远超过AMD。。根据JohnPeddieResearch的数据，NVIDIA占据了GPU市场约86%的份额。

不难理解，在生成式AI的火热浪潮下，英伟达被视为最大的潜在赢家。。从财报来看，这一波生成式AI对英伟达的需求主要体现在数据中心业务上。事实上，在整个2023财年的四个季度中，数据中心已经取代英伟达开始的支柱业务——游戏，成为最大的业务。

2022财年第四季度，——2023财年第四季度，英伟达各个部门的营收情况如何？|截图来源：英伟达

2023财年，数据中心总收入增长41%。，达到创纪录的150.1亿美元。仅第四季度，数据中心营收为36.2亿美元，约占英伟达&#039；总收入。

数据中心增长的基本盘来自新一代旗舰产品H100出货量的持续增长，云渗透率的持续提升，以及超大规模客户对AI布局的拓展。

就H100而言，在第二季度，它的收入已经远远高于A100，而后者&#039；的收入份额继续下降。据悉，H100在训练上比A100快9倍，在基于Transformer的大型语言模型推理上比A100快30倍。。

与此同时，NVIDIA正在为越来越多快速增长的云服务提供商(CSP)提供服务。，包括甲骨文和一些专注于GPU的云服务提供商(GPU专门化的CSP)。在过去的四个季度中，CSP客户贡献了大约40%的数据中心收入。

02下一步：AI即服务

在财报电话会议上，老黄透露了Nvidia&#039s新趋势：——AI企业服务走向云端。尽管更多的信息将会在十天后的GTC会议上公布。然而，英伟达正在与领先的云服务提供商合作，提供人工智能即服务，以便企业可以访问英伟达&#039；sAI平台。根据官方消息，客户将能够使用英伟达AI的每一层(包括AI超级计算机、加速库软件或预先训练的生成式AI模型等。)作为云服务。

老黄阐述，&quot技术突破的积累让AI走到了一个拐点。。生成式人工智能的多功能性和能力引发了全球企业开发和部署人工智能战略的紧迫感。然而，AI超级计算机基础设施、模型算法、数据处理和训练技术仍然是大多数人无法克服的障碍。

基于这样的行业痛点NVIDIA的下一个级别&#039；的商业模式是帮助每一个企业客户使用AI。

客户可以使用自己的浏览器，通过英伟达DGX云使用英伟达DGXAI超级计算机。，已经可以在甲骨文云基础设施上使用，预计将在不久的将来在微软Azure、谷歌云等平台上推出。在人工智能平台的软件层，客户将能够访问英伟达人工智能企业，以培训和部署大型语言模型或其他人工智能工作负载。在人工智能模型中，服务层英伟达将向希望为其业务构建专有生成式人工智能模型和服务的企业客户提供定制的NeMo和BioNeMo人工智能模型。从市场前景来看，黄仁勋认为ChatGPT让人

们意识到计算机编程的民主化，几乎任何人都可以用人类的语言向一台机器解释一个特定的任务。因此，世界上人工智能基础设施的数量将会增加。你会到处看到这些人工智能工厂&quot。人工智能的生产会像制造业一样。未来，几乎每个公司都会以智能的形式生产软件产品。数据进来，只做一件事，用这些数据生成新的更新模型。

他进一步解释了AI工厂，&quot当原材料进入时，建筑或者基础设施会启动，然后出现一些改进的东西，这是很有价值的。这就是所谓的工厂。所以我希望看到全世界的AI工厂。其中一些将被托管在云中。其中一些将是本地的。有些会很大，有些会很大。然后还会有一些小一点的。所以我完全期待这一天的到来。

其实老黄&#039；的人工智能工厂愿景正在发生。上个月，他在一次公开演讲中声称，自从ChatGPT出现以来，大约500家新的创业公司可能已经开发出令人愉快和有用的人工智能应用程序。

基于这样的前景，Nvidia对数据中心的未来充满信心。首席财务官克雷斯波说通过新的产品周期、生成式AI以及人工智能在各行业的不断采用，数据中心部门将继续实现增长。她说：&quot除了与每一个主要的超大规模云服务提供商合作，我们还与许多消费互联网公司、企业和初创企业合作。。这个机会意义重大，促进了数据中心的强劲增长，今年还会加速。

03汽车涨游戏跌

除了数据中心，NVIDIA&#039s其他业务板块有——游戏、汽车、职业视觉等等。，本季度&#039；的表现喜忧参半。

其中，汽车业务表现亮眼。本财年总收入增长60%，达到创纪录的9.03亿美元。第四季度营收创下2.94亿美元的纪录，比去年同期增长135%。，比上一季度增长17%。

车辆业务持续增长，无论是环比还是同比。英伟达称，这些增长反映了自动驾驶解决方案销售的增长，电动汽车制造商计算解决方案和人工智能驾驶舱解决方案的强劲销售。。电动汽车和传统OEM客户的新项目推动了这一增长。

值得注意的是，在今年1月初举行的CES大会上，英伟达宣布与富士康建立战略合作伙伴关系。共同开发基于英伟达DRIVEOrin和DRIVEHyperion的自动驾驶汽车平台。

相比之下，游戏业务依然深陷泥潭。在过去的几个季度里，RTX4080的销售疲软，

电子游戏行业的衰落，加密货币市场的疲软，去库存的压力让英伟达&#039；美国游戏业务陷入低迷，尤其是在第三季度，游戏业务收入同比暴跌51%。但正如首席财务官克里斯所说，最低点可能已经过去，情况可以改善。

第四季度，Nvidia&#039s游戏营收18.3亿美元，同比下降46%，环比增长16%，整个财年营收下降27%。。本季度和财年的同比下降反映了销售额的减少，这是全球宏观经济低迷和中国&#039；疫情控制自由化对游戏需求的影响。

不过，与第三季度相比，NVIDIA&#039美国游戏业务取得了一定的增长。。这得益于基于阿达洛芙莱斯架构的全新GeForceRTXGPU的推广。黄仁勋也肯定了这一观点，他说：游戏行业正在从新冠肺炎疫情后的低迷中复苏。而玩家热烈欢迎采用AI神经渲染的Ada架构GPU。

最近，游戏行业复苏的一个好迹象是，动视暴雪在第四季度实现了正收入增长。，超出预期。不过还是要警惕3354动视暴雪在PC和主机上销售游戏，而只有PC销售和英伟达有关，主机厂商使用AMD显卡。

另外，在财报发布的前一天，Nvidia宣布与微软签署了一项为期10年的协议，将XboxPC游戏阵容引入GeForceNOW。，包括《我的世界（Minecraft）》，《光环（Halo）》和《微软模拟飞行（MicrosoftFlightSimulator）》。微软完成对动视的收购后，GeForceNOW会增加《使命召唤（CallofDuty）》、《守望先锋（Overwatch）》等游戏。

除了游戏业务，专业视觉和代工的业务也比上年大幅下降。可以看出，半导体市场正在经历一个罕见的下行周期。

第四季度专业视觉业务收入为2.26亿美元，与去年同期相比下降65%。，比上一季度增长13%。本财年总收入下降27%，至15.4亿美元。本季度和本财年的同比下降反映了为帮助减少渠道库存而向合作伙伴销售的减少。链式增长是由桌面工作站GPU推动的。

OEM及其他收入同比下降56%，环比增长15%。财年收入下降了61%。本季度和本财年的同比下滑是由笔记本OEM和加密货币挖掘处理器(CMP)推动的。2023财年，，CMP营收可以忽略不计，2022财年为5.5亿美元。

04为什么赢家在风中？为什么是英伟达？

英伟达30年的发展历史可以分为两部分。从1993年到2006年。英伟达&#039；的目标是在竞争激烈的显卡市场中生存下来，并创造GPU的革命性技术。2006年到2

023年的转型，主要是关于如何利用CUDA这个平台。将GPU应用于机器学习、深度学习、云计算等领域。

后者让英伟达走上了人工智能的征程。今天它的市值已经超过了老牌霸主英特尔和AMD，这也是Nvidia在今天之下再次站在风口浪尖的前提&#039；s的生成式AI热潮。

在2019年的主题演讲中，黄仁勋分享了英伟达的起源&#039；s对行业的反复追查，发现了真正重要的问题并坚持下来。他说：&quot这使我们能够一次又一次地发明、重塑我们的公司和追溯我们的行业。。我们发明了GPU。我们发明了程序着色。我们把电子游戏做得如此美丽。我们发明了CUDA，把GPU变成了虚拟现实的模拟器。

回到NVIDIA的起点。那时候Windows3.1刚刚出来，个人电脑革命刚刚开始。英伟达想找到一种方法，让3D图形消费化、大众化，让大量的人能够接触到这种技术，从而创造出一种当时并不存在的全新的行业——视频游戏。。他们认为，如果它被制造出来，它可能会成为世界上最重要的技术公司之一。

原因是三维图形主要表现的是对现实的模拟，对世界的模拟相当复杂。如果你知道如何创造一个真假难辨的虚拟现实，模拟物理定律，将人工智能引入万物，这一定是世界上最大的计算挑战之一。一路走来衍生出来的技术可以解决惊人的问题。

最具代表性的案例正是通过CUDA和其他程序，它给计算和人工智能带来了创新性的影响，也使其在这一波生成式AI中处于最佳的位置。

虽然GPU作为一种计算设备的发现通常被认为有助于引领&quot寒武纪大爆炸&quot围绕深度学习，GPU并不是孤军奋战。英伟达内外的专家都强调，如果英伟达没有在2006年将CUDA计算平台加入投资组合，深度学习革命就不会发生。

CUDA(计算统一设备架构)计算平台是NVIDIA在2006年推出的软件和中间件堆栈。其通用并行计算架构使GPU能够解决复杂的计算问题。。通过CUDA，研究人员可以编程和访问GPU实现的计算能力和极端并行性。

在NVIDIA发布CUDA之前，给GPU编程是一个漫长而艰苦的编码过程，需要大量的低级机器码。。有了免费的CUDA，研究人员可以在NVIDIA硬件上更快更便宜地开发他们的深度学习模型。

CUDA的发明源于可编程GPU的思想。英伟达认为，为了创造一个更美好的世界。首先要做的是先模拟出来，而这些物理规律的模拟是超级计算机负责的问题，是科

学运算的问题。所以，关键在于如何把超级计算机能解决的问题缩小到普通计算机的大小，让你先模拟出来。，然后生成图片。这使得英伟达走向可编程GPU，这是一个巨大的赌注。

当时NVIDIA花了三四年开发CUDA，才发现所有产品的成本都要翻近一倍。当时不能给客户带来价值，客户明显不愿意买单。

要想被市场接受，NVIDIA只能增加成本，不能涨价。黄仁勋认为这是一个计算架构的问题为了让开发人员对这种架构感兴趣，每台计算机都必须能够运行。所以他继续坚持，最终创造了CUDA。然而，在那段时间里，英伟达&#039；的利润&quot破坏性地&quot下降，其股票跌至1.50美元，在大约五年的时间里一直处于低迷状态。直到橡树岭国家实验室选择了英伟达&#039；的GPU来建造一台公共超级计算机。

然后，全世界的研究人员开始采用CUDA技术，一个又一个的应用，一个又一个的科学领域。CUDA已被用于不同的科学领域，如分子动力学、计算物理、天体物理、粒子物理和高能物理。两年前，诺贝尔物理学奖和化学奖的获得者也能够在CUDA的帮助下完成他们的研究。

当然，CUDA也为NVIDIA&#039s游戏，因为虚拟世界的流体力学和现实世界是一样的，比如粒子物理的爆炸，建筑物的倒塌，和英伟达在科学操作中观察到的是一样的，是基于同样的物理规律。

然而在CUDA发布后的前六年，Nvidia并没有&quot献身&quot到AI直到AlexNet神经网络的出现。在即将到来的GTC会议上，黄仁勋邀请了OpenAI联合创始人兼首席科学家IlyaSutskever，Sutskever见证了NVIDIA&#039人工智能领域的崛起。

Sutskever，AlexKrizhevsky和他的博士生导师GeoffreyHinton创立了AlexNet，这是计算机视觉领域的开创性神经网络。，在2012年10月赢得了ImageNet竞赛。获奖论文表明，该模型取得了前所未有的图像识别准确率。它直接导致了未来十年人工智能的主要成功故事，从谷歌照片、谷歌翻译和优步到Alexa和AlphaFold。

按照Hinton的说法，如果没有NVIDIA，AlexNet就不会出现。得益于数千个计算核心支持的并行处理能力，NVIDIA&#039sGPU已经被证明是运行深度学习算法的完美选择。。Hinton甚至在一次演讲中告诉在场的近千名研究人员，他们应该购买GPU，因为GPU将成为机器学习的未来。

在2016年接受福布斯采访时，黄仁勋说，他一直知道NVIDIA图形芯片的潜力不仅仅是驱动最新的视频游戏，但他没有&#039；不要指望转向深度学习。

事实上，英伟达的成功&#039；的深度神经网络GPU是&quot一个奇怪的幸运的巧合。"文章&quot硬件彩票&quot由一位名为SaraHooker的作者在2020年发表，讨论了各种硬件工具成功和失败的原因。

她说Nvidia&#039成功就像中了彩票。很大程度上，这取决于&quot硬件进展和建模进展之间的正确对齐时刻&quot。这种变化几乎是瞬间的。"一夜之间，13，000个CPU和两个GPU解决了这个问题。她说。"这是它的戏剧。。

然而，Nvidia不同意这种说法，并表示，自2000年代中期以来，Nvidia已经意识到GPU加速神经网络的潜力，尽管他们没有&#039；我不知道人工智能会成为最重要的市场。

Alexnet诞生几年后，Nvidia&#039美国客户开始购买大量GPU用于深度学习。那时，RobFergus(目前是DeepMind研究科学家)甚至告诉Nvidia应用深度学习研究副总裁BryanCatanzaro说有多少机器学习研究者花时间为GPU写内核？这太疯狂了。你真的应该研究一下。

黄仁勋逐渐意识到，人工智能是这家公司的未来，英伟达会立即将所有赌注押在人工智能上。

所以，在2014年的GTC主题演讲中，随着人工智能成为焦点，黄仁勋说机器学习是&quot高性能计算领域最激动人心的应用之一&quot。"取得激动人心的突破、巨大的突破、神奇的突破的领域之一，是一个叫做深度神经网络的领域。"黄仁勋在会上说。

此后，英伟达加速了AI技术的布局，不再仅仅是一家GPU计算公司，逐渐建立起强大的生态系统，包括芯片、相关硬件以及一套针对其芯片和系统优化的软件和开发系统。。这些最好的硬件和软件平台可以最有效地生成AI。

可以说GPUCUDA改变了AI的游戏规则。。中信证券分析师徐英博在播客节目中评论：英伟达一直在做一件非常聪明的事情，那就是软硬件一体化。在GPU硬件半导体的基础上，衍生出通用计算的CUDA。。这促成了英伟达&#039；软件和硬件的双重规模效应。

硬件方面，因为是图形和计算的统一架构，所以通用性保证了它的规模，规模摊薄了它的研发成本。所以硬件本身可以通过规模化获得研发成本的比较优势。

在软件方面，因为它有一个庞大的开发者&#039；s生态，还有这些珍贵的软件开发者，即使他们换了公司。但他可能仍在使用CUDA&#039s软件。

主要参考文献：

1)《ChatGPT火了，英伟达笑了》——中国电子报

Nvidia:GPUCompany(1993-2006)Nvidia:MachineLearningCompany(2006-2022)

4)conferenceonartificialintelligenceheldbyHuangRenxun,CEOofNVIDIAinMSOE

5)HuangRenxun'squestionandanswer:WhyMoore'slawisdead,butthemeta-universewillstillhappen

.[XY001]6)HowdoesNVIDIAdominateartificialintelligence,andplanstokeepthiswaywhenreproductiveartificialintelligencebreaksout

7)中信证券徐英博：从英伟达——小宇宙——创业内幕看国产GPU的挑战与前景