科技周期探索之八

AI 时代的三个案例公司：微软、AMD、英伟达

中性

核心观点

微软：AI 应用的领跑者

2000 年后的微软经历了十四年的鲍尔默时代，虽然这一时期微软的收入翻了 3.8 倍，利润翻了 2.3 倍，但由于错失了移动互联网卡位，其估值从 64 倍回到了最低仅有 10 倍。纳德拉接任后，一是打破了 Windows 的围墙花园，使微软变得更加开放；二是加码云计算，将云计算打造成为微软的新增长点；三是押注了 AI 的投资，尤其是在 OPENAI 上的大比例持股使得微软第一时间释放了大模型的能力，推广出一系列的 AI 新产品与服务，赢得先发优势。

从过去 10 年股价翻了 10 倍来看，公司已经获得了转型的成功，未来自研+投资 OPENAI 两条腿走路为其提供了更强的保障。

AMD：轻装出发的挑战者

AMD 从创始人桑德斯退休后，经历了四任 CEO：鲁伊兹收购了 ATI，为后续 AMD 进军 GPU 领域埋下了伏笔，但也因为收购价格过高使得 AMD 陷入巨大的财务压力；梅耶时代剥离了格罗方德，这让 AMD 变成了轻资产公司；里德时代则是面向低功耗与游戏市场转型，虽然务实但无法被股东理解；终于苏姿丰的到来将公司带入到再次腾飞的局面中。

在苏姿丰时代，定位高性能计算提升了 AMD 的品牌，同时 Zen 架构的出现以及台积电的代工，让 AMD 在 CPU 市场对英特尔竞争开始占优，加之服务器市场的成功，市场份额被一步步夺回。有了 CPU 的成功案例，公司开始通过研发与收购，快速布局到 GPU 市场上，2024 年 AI 芯片收入预计超过 50 亿美元。虽然进步很快，但由于英伟达的优势过于明显，因此市场对 AMD 在 AI 芯片市场未来份额能否达到像 CPU 那样高显然还有存疑，仍需要更多的观察。

英伟达：摩尔定律的延续者

在互联网泡沫时期就存在，并还由创始人执掌的公司，英伟达绝对是重要的一个。公司 2006 年开发出 CUDA，对手看不懂，股东也不理解，但最终 CUDA 的巨大成功带给公司的回报不仅是收入，而是将英伟达定位成了接替英特尔成为“摩尔定律”的延续者。

正是因为有着“指数型思维”，英伟达此后的种种工作都围绕着最大化地推动算力进步为目标：从两年一代的 GPU 架构，到 NVSwitch、NVLink 综合提升全栈能力，其在 GPU 的市场份额始终保持在 80%以上，留给竞争对手的机会少之又少。

在 AGI 到来之前，英伟达依然充满了机会，但期间若遭遇经济周期下行，其“靠客户竞争以获得产品溢价”的局面可能变化，可能带来毛利率的短期波动。

风险提示：地缘政治的不确定性，美联储降息幅度的不确定性，部分行业竞争格局的不确定性。

行业研究·海外市场专题

美股

中性·维持

证券分析师：王学恒
010-88005382
wangxueh@guosen.com.cn
S0980514030002

市场走势

资料来源：Wind、国信证券经济研究所整理

相关研究报告

《美元债双周报（25 年第 2 周）-非农数据强势，美债利率临近 4.8%》 ——2025-01-13
《美股市场速览-加仓窗口或将在近期出现》 ——2025-01-12
《美股市场速览-特斯拉有所回撤，小盘风格较优》 ——2025-01-05
《美元债双周报（24 年第 53 周）-美联储鹰派降息，美债利率高位徘徊》 ——2024-12-30
《美股市场速览-算力芯片巨头表现强劲》 ——2024-12-29

内容目录

图表目录

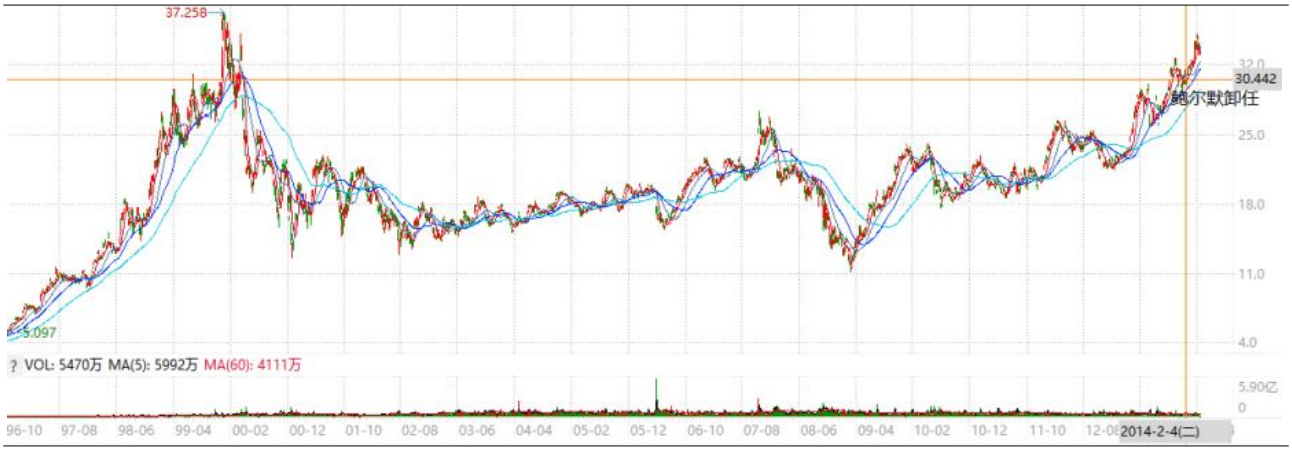
微软：AI 应用的领跑者

鲍尔默时期（2000-2014 年）：失去的移动互联网

华尔街评价一个公司的 CEO 优秀与否并不是收入和利润，而是股价。

2000 年 1 月 13 日，鲍尔默正式被任命为首席执行官，直到 2014 年 2 月 4 日正式卸任，在 14 年的时间里，微软公司市值始终无法逾越科网泡沫的高点。这使得市场逐渐对鲍尔默领导下的微软失去了耐心。鲍尔默甚至被 BBC 评为 2013 年最差 CEO 之一。2012 年 5 月，亚当·哈通在《福布斯》杂志的专栏中将鲍尔默描述为“美国大型上市公司中最糟糕的首席执行官”，并表示他“将微软赶出了增长最快、利润最丰厚的科技市场（移动音乐、耳机和平板电脑）”。

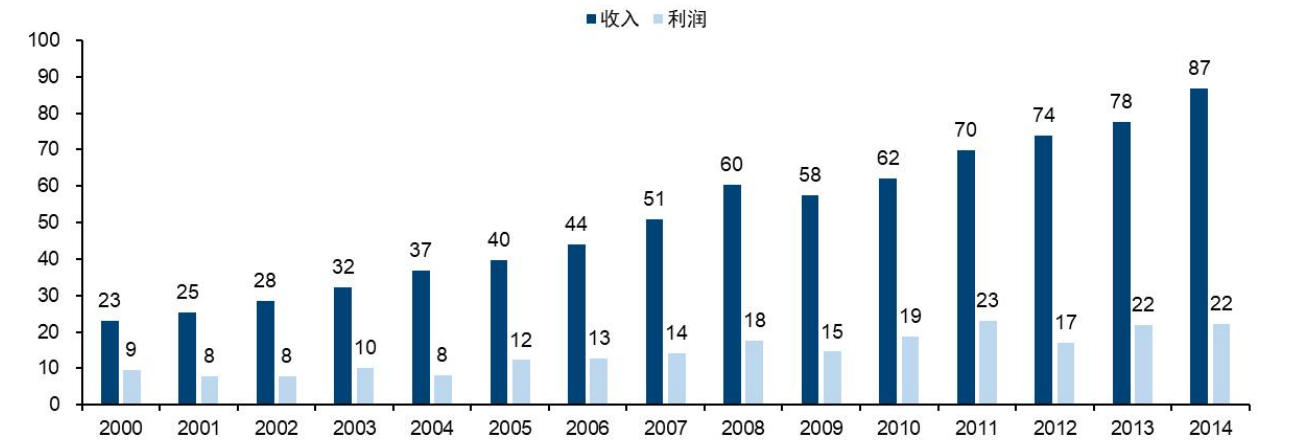
图1：鲍尔默直到卸任 CEO，微软公司的股价也没有能够创新高



资料来源：wind，国信证券经济研究所整理

鲍尔默时期，微软公司的收入翻了 3.8 倍，相当于年化增速 10%；利润翻了 2.3 倍，年化增速 6%。这个成绩要好于通用电气的韦尔奇和 IBM 的郭士纳。

图2：鲍尔默时期微软公司的收入与利润（十亿美元）



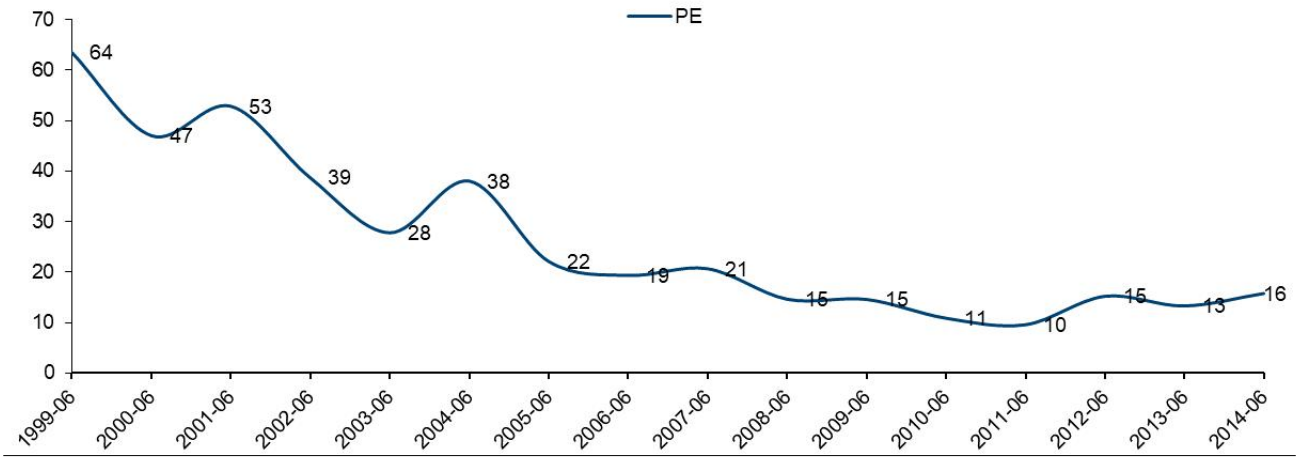
资料来源：Factset，国信证券经济研究所整理

但问题在于，这一期间微软几乎在与苹果、谷歌的较量中，完美地输掉了移动互联网的战役。我们曾在《2002-2016 年：移动互联网的大时代》中介绍，Windows

Mobile 是微软在 1996 年发布的手机操作系统，当时命名 Windows CE，后来经过了几个版本的迭代与改名，于 2003 年正式命名为 Windows Mobile。这要比乔布斯开发 iPhone，以及谷歌开发安卓要早得多，但最终因为产品设计力不够，体验不好而与移动互联网大潮失之交臂。

一边是苹果的新 Mac、iPod、iPhone、iPad、App Store 在消费电子领域、移动互联网领域高歌猛进，一边是 Windows Mobile 手机份额被蚕食，加之鲍尔默时期的 Surface 笔记本尚未有今天的影响力，因此微软公司让投资者看不到更大的希望在哪里。这导致了其市盈率从 1999 年的 64 倍左右跌落至 2011 年的 10 倍——10 倍 PE 的科技股是不多见的，除非投资人已经对公司不抱有成长的希望。

图3：鲍尔默时期的微软公司市盈率



资料来源：Factset，国信证券经济研究所整理

鲍尔默与盖茨是哈佛大学的同学，他的专业是数学与经济学，盖茨在成立微软不久，认为公司需要一个人去专心运营商务，于是 1980 年他说服了鲍尔默加入了微软。2000 年鲍尔默接替盖茨管理微软，从资历上讲，他领到过运营、操作系统开发以及销售和支持，已经是微软的绝对资深员工与实际的二号人物。虽然盖茨对软件行业的发展直觉和敏锐度是一流的，但微软在硬件上始终缺乏突破，此时的盖茨也好，鲍尔默也好，都缺乏硬件产品化的成功案例。从鲍尔默的履历上来看，他擅长的是营销与运营，而非产品设计与开发。我们曾在《案例篇：移动互联网的十倍股和百倍股》中提及，虽说成功公司的路径各有不同，但公司的创始人的产品能力是卓越的，例如苹果的乔布斯，亚马逊的贝佐斯，奈飞的哈斯汀斯，腾讯的马化腾，脸书的扎克伯格，谷歌的佩奇和布林，Salesforce 的贝尼奥夫。因此当鲍尔默遇到同时期的乔布斯与佩奇和布林，其在产品上的前瞻力注定略逊一筹，这也算反向佐证了“成功的科技企业必然需要一个优秀的产品带头人”。

鲍尔默执政期间，微软斥巨资收购的雅虎与诺基亚，回头看都属于创新浪潮前的平台，并没有给公司带来质的改变。反倒是微软 2007 年投资了脸书获得了不错的投资收益，但投资比例又过低，仅占 1.6%。

纳德拉早期（2014-2018 年）：打破围墙花园

印度裔高管萨蒂亚·纳德拉 2014 年接替史蒂夫·鲍尔默担任 CEO 成为了微软的一个转折点。纳德拉是电气工程学士和计算机硕士，毕业先在 SUN 公司工作两年，并于 1992 年加入微软。

他曾历任商业解决方案副总裁（2001-2006）、在线服务高级副总裁（2007-2011）、服务器和工具部门总裁（2011-2014）。从纳德拉的履历中可以看出，他是典型的工科背景，且在微软的所有经历都是产品端，因此其更了解产品。同时，他在 CEO 之前的最后一个岗位是 Azure 云平台的总裁，这也是他未来要加码的方向，或者说这恰恰是盖茨与鲍尔默选中纳德拉的原因。

表1：纳德拉的简历

时间	职务	主要经历
2014-	CEO	领导微软向云端优先、移动优先的转型，并监督了多家公司的收购，包括 LinkedIn 和 GitHub。
2011-2014	服务器和工具总裁	负责管理 Azure 云平台以及适用于公司数据中心的产品，Windows server 和 SQL server 数据库。
2007-2011	微软在线服务高级副总裁	负责 Bing、微软 Office、Xbox live 和其他商业软件。
2001-2006	微软商业解决方案副总裁	微软商务平台的开发，包括微软 Commerce Server 和微软 BizTalk Server，Great Plains，Dynamics。
2000-2001	bCentral 副总裁	面向小型企业的网络服务，包括托管网站和电子邮件。
1992-2000	工程师	从 SUN 公司跳槽加入微软，参与的首批项目包括不成功的互动电视产品和 Windows NT 操作系统。
1990-1992	工程师	毕业后第一份工作在 SUN 公司

资料来源：微软，国信证券经济研究所整理

作为公司的掌舵人，光有产品的敏锐度是不够的，战略方向感是更为重要的。在纳德拉 2017 年的个人新书《刷新：重新发现微软灵魂并为每个人畅想更美好未来的探索》中，纳德拉提及他最看好的三个方向：混合现实、量子计算、人工智能。事实证明，他正是领导微软公司在一步步地向着这些目标迈进。

2014 年 2 月，我的世界（Minecraft）注册用户达到 1 亿，其中许多用户都是儿童，代表了最新一代的游戏玩家和软件用户。2014 年 9 月，斥资 25 亿美元收购 Mojang（Minecraft 的开发公司）为微软的产品提供了庞大的潜在客户群，这是微软自纳德拉上任后的首次重大收购。微软将该游戏扩展到 Xbox 之外，增加了新功能和内容，并推出了专为课堂使用的教育版“我的世界”。为了展示该游戏的多种潜在用途，微软在其华盛顿州雷德蒙德园区内推出了一个“我的世界”版本，以便员工可以了解该设施的升级情况。

图4：我的世界（Minecraft）

资料来源：Minecraft.net，国信证券经济研究所整理

图5：雷德蒙德园区（我的世界版）

资料来源：微软，国信证券经济研究所整理

2015 年 7 月，微软推出了 Windows 10，这是其桌面操作系统 Windows 8.1 的后续版本。它带来了新功能，包括 Cortana 智能助手、Edge 浏览器、Xbox 游戏流媒体功能、新的生物识别登录选项以及在平板电脑或智能手机模式与桌面模式之间切换。Windows 10 推出四周就已覆盖 7500 万台设备，超过了微软之前的所有发布版本。

2016 年 9 月，微软人工智能研究院成立，它汇集了 5000 余名计算机科学家及工程师，拓展公司的人工智能领域力量。团队由沈向阳（Harry Shum）领导，还包括信息平台团队、Cortana 和 Bing 团队，以及环境计算和机器人团队。这标志了在纳德拉接任微软 2 年之后，其在 AI 方向加大投资的决心已经得到了董事会的认可。2017 年 9 月，盖茨在评价纳德拉时提及：“他正在对人工智能和云计算等几项关键技术进行大举投资，微软将在这些技术上脱颖而出。”

2016 年 12 月，微软斥资 262 亿美元收购 LinkedIn，目标是发展专业社交网站，并将其与微软的企业软件整合在一起。这次收购让微软能够接触到 LinkedIn 庞大的用户群。

由于在我们的划分中，2016 年是移动互联网周期的结束，也是人工智能时代的开始，移动互联网中最大的杀手级应用莫过于社交，类似于微信、Facebook/WhatsApp，由于谷歌收购了 YouTube 且有了安卓系统，苹果则是占据了 iOS 并形成强大 App Store，微软收购 LinkedIn 总算是“赶了个晚集”，弥补了遗憾。2016 年微软收购 LinkedIn 时，它的用户数不足 5 亿，2024 年用户数已经突破 10 亿。收入上，2017 财年 LinkedIn 收入仅为 22.7 亿美元，到了 2024 财年收入则增长至 163 亿美元，年化复合增速为 14.3%，这还不算其带来了巨大的网络效应与整合效应，因此这是一笔相当成功的收购，尤其比起来鲍尔默时期对雅虎与诺基亚的收购。

图6：Windows 10（2015）

资料来源：微软，国信证券经济研究所整理

图7：2016 年微软收购领英

资料来源：微软，国信证券经济研究所整理

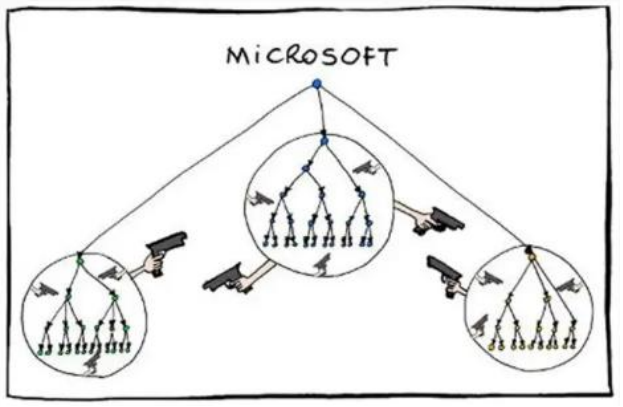
2018 年 6 月，微软宣布以 75 亿美元的价格收购 GitHub。GitHub 是一个在线软件源代码托管服务平台，用于公开程序或软件的代码。微软的收购采取了包容的态度，GitHub 与 LinkedIn 类似，继续作为社区，平台和业务独立运作。GitHub 的蓬勃发展并不是因为微软大力营销和销售它，而是因为微软坚持 GitHub 保留其开源精神和开发者至上的文化。

虽然当时从财务的角度，这笔收购显得过于“慷慨”，但对于微软而言，接触每天使用 GitHub 代码库产品的大量开发者，这样他们就可以被引导到微软的开发者环境中，真正的意义是在这里。纳德拉在 2021 年曾表示：“我们提供最受欢迎的工具，帮助开发人员快速从创意到代码，再从代码到云。Visual Studio 每月拥有超过 2500 万活跃用户，GitHub 拥有近 6500 万开发人员，在过去的 12 个月中，使用 GitHub 的月活跃组织数量增加了 70%。”

图8：2018 年微软收购 GitHub

资料来源： 微软，国信证券经济研究所整理

图9：消除部门之间的争斗是纳德拉的目标



资料来源：bonkersworld.net，国信证券经济研究所整理

纳德拉接任微软时，微软最大的诟病是其产品的封闭性。众所周知，微软捍卫其专有的 Windows 和 Office 软件，并谴责开源替代品。2011 年以讽刺科技漫画闻名的设计师马努·科内特发布了在微软内部的山头林立，互相敌对的漫画。

纳德拉上任则赋予了微软新的目标：微软的存在是为了“让地球上的每个人和每个组织都能取得更大的成就”，这寓意着微软将成为一家以人为本的公司，而不是一家产品公司。微软首席营销官克里斯·卡波塞拉表示：“我们从一种‘无所不知’的文化转变为一种‘无所不学’的文化。我们现在所做的一切都植根于成长型思维。”这种强调“同理心”，提出要对员工和客户保持开放与尊重，改善企业文化和组织架构，消除内部的隔阂，鼓励协作与沟通，逐渐成了公司的主旋律。

2014 年 3 月，微软推出的 Office for iPad 为用户带来了 Word 和 PowerPoint 等跨设备应用。此次发布包括适用于 iPhone 的 Office 新功能和适用于 iPad 的更新应用，随后不久又推出了适用于 Android 的 Office。纳德拉在担任 CEO 后的首次公开演讲中宣称，“云计算和移动的神奇结合”，并表示微软“绝对致力于让我们的应用程序跨平台运行”。跨平台开发应用，虽然是 Windows 前进的一小步，但是却是公司从封闭走向合作与开放的开始。

以下是 Windows 逐步走向开放的一些案例：

2017 年，微软与 Dell 合作推出基于 Windows 的 Dell Latitude 系列笔记本电脑；

2017 年，微软与 Amazon 合作推出基于 Azure 的 AWS 云服务；

2018 年，微软宣布开放 Windows API，允许开发者使用 Windows API 开发应用程序；

2018 年，微软宣布与 Box 合作推出基于 Azure 的 Box 云存储服务；

2018 年，微软宣布与 VMware 合作推出基于 Azure 的 VMware 云服务；

2019 年，微软宣布与甲骨文合作推出基于 Azure 的甲骨文云服务；

2019 年，微软宣布与 Google 合作推出基于 Chrome OS 的 Windows 应用程序，该合作允许用户在 Chrome OS 上运行 Windows 应用程序，扩大了 Windows 的应用场景；

2019 年，微软宣布与 Linux 基金会合作推出基于 Linux 的 Windows Subsystem for

Linux（WSL）。WSL 允许用户在 Windows 上运行 Linux 应用程序，提高了 Windows 的灵活性和可扩展性；

2019 年，微软将 Windows Calculator 开源，允许开发者在 GitHub 上贡献代码和参与开发，这是微软首次将 Windows 组件开源。

加码云计算（2018-2022 年）：Azure 乘风起

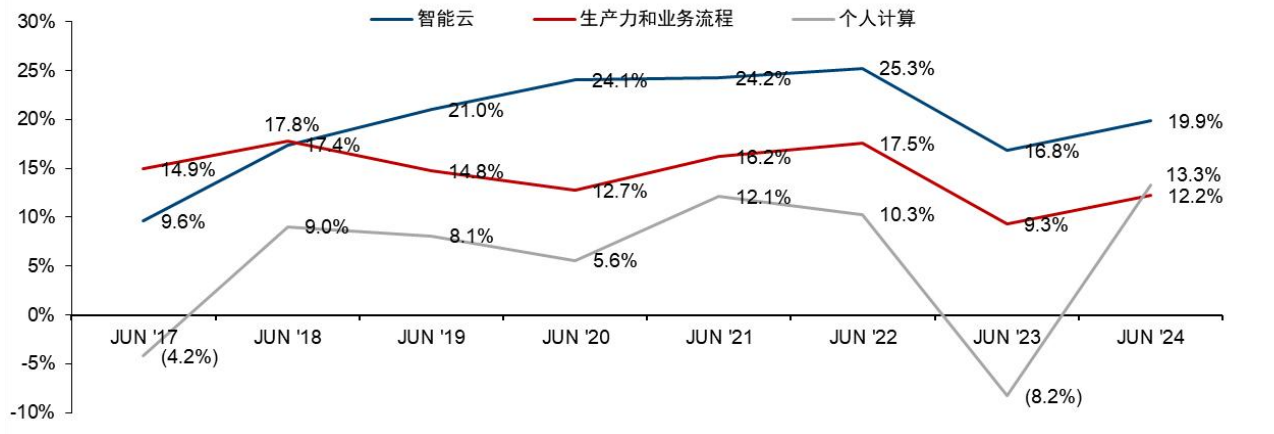
在微软财报中，有三部分内容：

智能云（Intelligent Cloud）：包括 Azure（微软的云计算平台，提供计算、存储、数据库、安全、网络等服务）、Azure Stack（一个混合云平台，允许客户在自己的数据中心或云端环境中运行 Azure 服务）、Microsoft Azure AI（基于云的人工智能平台（基于云的物联网平台，提供设备管理、数据分析、安全等服务），提供机器学习、自然语言处理、计算机视觉等服务）、Microsoft Azure IoT、Power Apps（低代码开发平台，允许用户创建自定义的商业应用程序）、Power Automate（自动化平台，允许用户自动化商业流程和任务）、Microsoft 365 Security（安全解决方案，提供身份验证、访问控制、威胁防护等服务）、Microsoft 365 Compliance（合规性解决方案，提供数据保护、隐私保护、法规遵从等服务）；

生产力和商业流程（Productivity & Business Processes）：包括 Office 软件、Microsoft 365、Dynamics 365（ERP 和 CRM）、LinkedIn、Skype for Business、Microsoft Teams（团队协作平台）、OneDrive（云存储）、Outlook；

个人计算（More Personal Computing）：包括 Windows、Surface 电脑、游戏（Xbox 游戏机、Xbox 游戏软件和服务、Xbox Live 在线游戏服务）、搜索与广告（包括 Bing 搜索引擎和 Microsoft Advertising 广告服务等业务）、应用商店。

图10：2017-2024 财年，微软三大业务板块的收入增速



资料来源：Factset，国信证券经济研究所整理

其中，支撑微软在 2018 年以后增长最快的是云计算，其核心是 Microsoft Azure。Azure 平台于 2010 年 2 月正式推出，当时的名称为 Azure Service Platform，包含 Azure 云计算、Azure 存储、SQL Azure 与 AppFabric 四种服务，且仅提供 PaaS。2011 年纳德拉成为了服务器和工具部门的总裁之后，Azure 在 2012 年进步很快：更新管理接口，采用 HTML5 技术；发行 IaaS，包含虚拟机与虚拟网络；发行 Website 服务，并首次支持.NET 以外的平台；发行 Media Service 服务。到了 2014 年下半年 Azure 发行了 Mobile Service，提供移动应用必须的后台服务，包含资料、

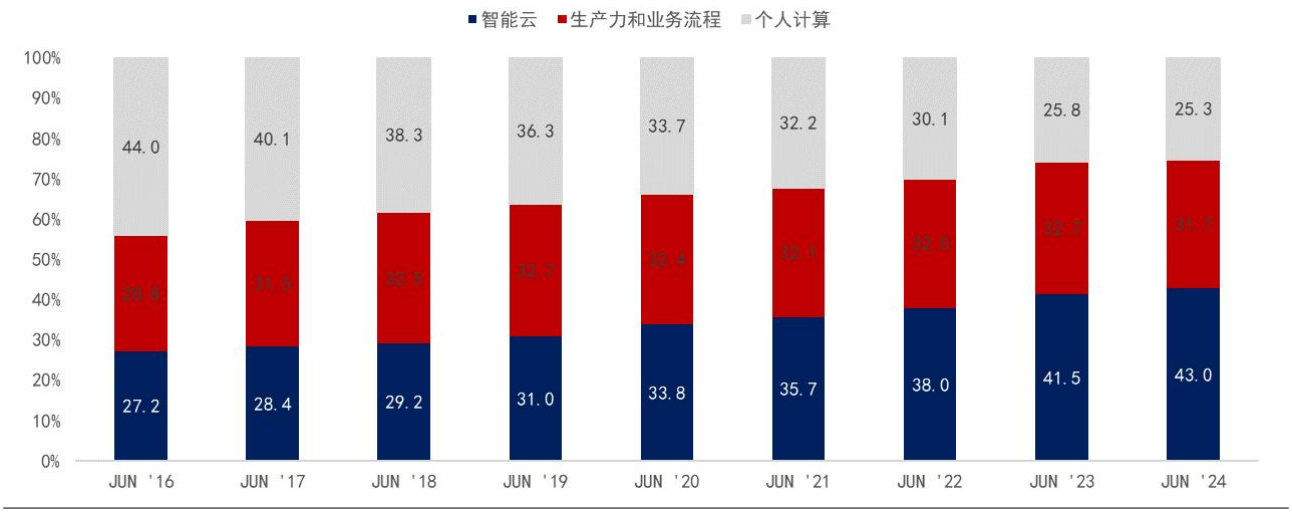
识别、通知以及 API 等。

2014 年纳德拉成为 CEO 当年，Windows Azure 更名为 Microsoft Azure，以修正其市场方向，也为了让外界不再认为 Azure 只能运行 Windows 操作系统；2015 年微软将 Website 与 Mobile Service 合并，并新增 API App 与 Logic App 合称为 Azure App Services；还推出了 Azure Application Insights 以支持应用程序层级的监测数据能力；新增 Azure DNS 以支持 DNS 托管、Azure Search 支持搜索能力等。2016 年微软推出 Azure Functions(函数服务)以支持无服务器(Serverless)的应用，成为继 AWS Lambda 与 Google CloudFunction 之后的第三个具备无服务器应用程序开发能力的主流云供应商，同时也推出了 Service Fabric 以支持微服务（Microservices）的开发。

由于 2016 年之后云计算行业处在大发展时期，加之纳德拉很清楚微软不能错过这个巨大的机会，因此微软也大踏步投入云计算。除了增加基础设施，增加应用功能，前述开放策略也为微软赢得了多个高质量合作伙伴。到了 2018 年，微软宣布了 54 个 Azure 区域，比任何其他云提供商都多，服务范围覆盖全球 140 个国家。此次扩张巩固了该公司作为云计算全球领导者的地位。微软增加了近 500 项新的 Azure 功能，推出了首创的混合和云到边缘解决方案 Azure Stack（混合云）和 Azure Sphere（物联网），并达成了创纪录的数百万美元的商业云协议。

云计算在 2016 年占收比为 27.2%，是最小一个业务板块，到了 2024 年占收比来到了 43%，成为最大的业务板块。2018 年-2022 年，云计算业务连续四年加速增长。

图11：2016-2024 财年，微软三大业务板块的占收比



资料来源：Factset，国信证券经济研究所整理

Azure 和 AWS 都是云计算巨头，各有其优势。但 Azure 作为后发者，其在几个方面上表现出一定的优势，包括：

混合云策略：Azure 提供了混合云策略，允许客户在自己的数据中心、Azure 云端和其他云平台之间进行选择和集成；

企业软件整合：Azure 提供了与 Microsoft 企业软件的紧密整合，例如 Office 365、Dynamics 365 等；

开发者体验：Azure 提供了友好的开发者体验，例如 Visual Studio、Azure DevOps

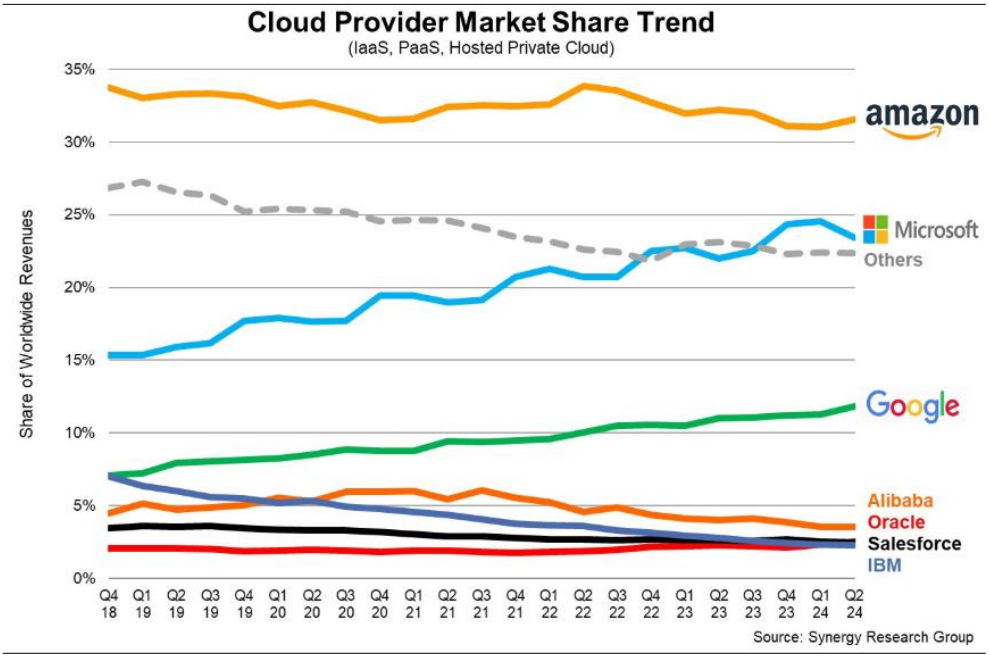
等；

AI 和机器学习：Azure 提供了基于 AI 和机器学习的云计算服务，例如 Azure Machine Learning、Azure Cognitive Services 等。

此外，Azure 在 2020 年发布了 Azure Arc，加强了混合云和边缘计算领域，同年发布了 Power Platform，加强了低代码开发和商业应用领域；2021 年收购了 Nuance Communications，标志着公司在人工智能和自然语言处理领域的扩展；2022 年发布了 Cloud for Sustainability，加强了可持续发展和环境保护领域。

总结下来，从 2018 年 Azure 的市场份额从 15%逐步攀升至 2023 年的 24%，这是纳德拉接管微软后最成功的例证。

图12：2018-2024 年云计算企业市场份额



资料来源：Synergy research，国信证券经济研究所整理

2023 年 10 月，微软也完成动视暴雪的收购，斥资共 687 亿美元，耗时 22 个月。由于动视暴雪是世界上顶尖的游戏公司，收购这么大规模的游戏公司要通过各国的反垄断审查，阻力较大的是美国、欧洲与英国。为了打消英国监管机构的顾虑，微软把动视暴雪持有的云游戏版权出售给法国的育碧（Ubisoft）。

收购动视暴雪是微软游戏整合趋势的一部分，它拥有的著名特许经营权包括《使命召唤》、《暗黑破坏神》和《魔兽世界》，收购动视暴雪后，微软成为全球收入第三大的游戏公司。游戏行业预计到 2030 年将增长至 5000 亿美元以上，收购动视暴雪可能会为微软带来丰厚利润。从短期来看，动视暴雪的游戏加入 Xbox 库和其他平台将为微软的游戏收入带来适度提升。

动视暴雪每月活跃用户数为 3.56 亿，尤其随着游戏从主机转向移动设备，动视暴雪的移动用户对微软至关重要。展望未来，由于游戏是元宇宙的一个天然特性，这和纳德拉心中三大方向“元宇宙、量子计算、人工智能”相匹配。

纳德拉评价收购动视暴雪时说：“当我想到动视暴雪的产品组合时，它为我们提供了覆盖 PC 和游戏机的大量资产，当然还有覆盖移动端，这是我们以前从未有过的。现在我们既有内容，又有能力访问人们玩游戏的所有传统高规模平台，即游

戏机、PC 和移动设备。”

从市场反馈来看，2022 年 1 月 18 日，微软宣布收购动视暴雪后，索尼公司股价当日下跌了 7.2%（日股则下跌了 12.8%），在后续，索尼公司的股价（美股）2 年后依然没有回到曾经的位置。这说明微软对动视暴雪的收购影响了行业竞争格局，这种预期进而影响了索尼的估值水平。

图13：索尼（SONY.N）股价

资料来源：wind，国信证券经济研究所整理

最后，按照微软对 LinkedIn 及 GitHub 的管理方式，动视暴雪也将获得相对较大的自主权，在收购协议中，微软承诺将尊重动视暴雪的文化和独立性，并允许其继续发展和运营自己的游戏业务。因此，动视暴雪将继续相对独立地经营自己的业务，同时也将受益于微软的资源和支持，这将使得动视暴雪能够更好地发展和扩展自己的游戏业务，并为玩家提供更多的游戏选择和体验。

AI 再出发（2023-2024 年）：押注 OpenAI

2022 年 11 月，OpenAI 公司的 ChatGPT 3.5 横空出世，这颠覆了人们对语言模型的认知。它凭借着如此流畅、丝滑的输出，开启了大模型（Transformer 架构）时代的新征途。

表2：OpenAI 的员工人数

年	员工人数	增长（数量）	同比（%）
2015	10	-	-
2016	25	15	150%
2017	45	20	80%
2018	80	35	77.8%
2019	150	70	87.5%
2020	250	100	66.7%
2021	300	50	20%
2022	375	75	25%
2023	770	395	105.3%
2024	3,531	2,761	358.6%

资料来源：seo.ai，国信证券经济研究所整理

OpenAI 公司成立于 2015 年 12 月，总部位于旧金山，它起步于非盈利组织，筹资主要是捐款，尽管其捐助人承诺捐款 10 亿美元，但是截至 2019 年，实际募集到的捐款总额仅为 1.3 亿美元。

在我们的报告《2016-2030 年：通用人工智能时代的到来》中曾介绍，2017 年 transformer 论文发布，到谷歌开发出 BERT 模型，OpenAI 也敏锐地发现出这是一条新路，并于 2018 年 6 月与 2019 年 2 月，研发了 ChatGPT 1.0 与 ChatGPT 2.0。随着对 LLM 的理解，他们发现只接受捐款的方式很难承受得起巨大的机器学习投入，因此在 2019 年，OpenAI 从非营利性组织转型为“有上限”的营利性组织。OpenAI 称有上限的利润模式使 OpenAI 能够合法地吸引风投，还可以向员工授予公司股份。

但问题是，这涉及到 OpenAI 的初衷，2015 年，作为非营利组织，OpenAI 的定位是：

1、OpenAI 是一家非营利性人工智能研究公司；

2、目标是以最有可能造福全人类的方式推进数字智能，不受产生财务回报需求的限制；

3、由于研究不受财务义务的限制，公司可以更好地专注于对人类的积极影响。

当时对 OpenAI 支持最大的马斯克是想打造出一个真正的“OPEN 的 AI”，以区别于谷歌的“封闭的 AI”，所以才取了这个名字。而当 OpenAI 真的打开了“潘多拉”魔盒，看到了通往 AGI 的巨大机会之后，它的诸多参与者，又希望将 OpenAI 变成一家盈利公司以获得更多的融资或者上市。

奥特曼加入的一年实际上就是从公司的从非营利化到盈利化转折的一年。但是 OpenAI 引入微软 130 亿美元的投资又是个权宜模式，这种非盈利组织再控股一个盈利实体的架构是较为特殊的，有诸多的约束。但如果直接转向营利组织，董事会又表示强烈的反对。

随着公司的不断扩大，这种矛盾越来越激烈，2023 年 11 月份，奥特曼先被董事会罢黜，之后微软则支持了奥特曼的回归。2024 年 10 月，OpenAI 融资 66 亿美元，投后估值 1570 亿美元，短短 9 个月时间公司估值接近翻倍，融资由 Thrive Capital 领投，微软继续参投，英伟达、软银都首次投资 OpenAI，其他投资方还包括 Khosla Ventures、Altimeter Capital、富达、Tiger Global、阿联酋投资公司 MGX 等。

2024 年 12 月，OPENAI 公司宣布重组，公司正式一分为二：一部分是非营利机构，另一部分转型为特拉华州公共利益公司（PBC）。OpenAI 是一个典型的在遇到巨大的发展机遇后发生了初衷改变的公司，于是无论从管理层还是公司的核心员工，都因为这种转变而受到影响，坚持公益初衷的员工逐渐离职，坚持公益初衷的董事会则反对盈利化转变，而期望资本助力的投资人与奥特曼则极力想实现盈利化的转变，这个对抗使得公司发生了本不必要的巨大内耗。

对于微软而言，在投资 OpenAI 后获益颇丰。

一来其协议保障了微软公司可以较快的收回成本，包括：1、与 OpenAI 形成独家合作伙伴关系，以开发和商业化 AI 技术，包括 OpenAI 的语言模型和其他 AI 工具；2、OpenAI 将其 AI 模型与微软的 Azure 云计算平台集成，使开发者更容易构建和部署 AI 驱动的应用程序；3、微软有权获得 OpenAI 高达 75%的利润，直到其收回 130 亿美元的投资；4、OpenAI 累计利润达到 920 亿美元之后，微软的分红比例下降，剩余部分利润由其他风险投资者和 OpenAI 的员工分享；5、当利润达到 1500 亿美元之后，微软和其他风险投资者的股权将无偿转让给 OpenAI 的非营利基金。

二来是在与 OpenAI 合作的过程中，微软既学习了 OpenAI 以及大模型的开发现状，

又推出了多种 AI 产品，包括 Microsoft Copilot 于 2023 年推出，明显拉动了微软各个业务条线的收入增速。微软通过将 OpenAI 集成到其产品和服务中，尤其是从其云业务中，云托管服务的需求支持 Azure OpenAI 与每家公司的应用程序和软件的集成。此外，CRM、会计和网络安全、Bing、ERP、Office、编程软件和操作系统中集成 OpenAI 模型方面比竞争对手更具优势。

悲观的人，经常提到一些问题：例如桌面 AI 代码质量问题、错误率仍高、安全性不高；云 AI 则缺乏广泛的语言支持、需要大量的数据和训练；行业 AI 则缺乏对行业应用的广泛支持，功能尚简单等等，但这些问题都将伴随大模型的优化，成本的降低，Agent 的数量增加而被改善。换句话说，这也是大模型未来的潜力与想象力之所在。

着眼未来，很大程度上的变化是，之于移动互联网的 IOS 与安卓，到底现在的大模型称之为“平台”，还是类似 Azure MaaS、Vertex AI、Meta AI、Bedrock 才称之为“平台”，倘若是在 LLM 层之上的平台越强，则他们可以集合更多的 LLM；倘若 LLM 是稀缺资源，则它们的强大更能促使各个平台向它们靠拢。例如，谷歌目前的思路是垂直一体化思路，即从应用到芯片都自己开发，希望做成 AI 时代的“苹果”；而其他公司要么摆脱不了英伟达的芯片能力，要么在除了 AI 平台之外的某个环节是开放的。

但有一点是肯定的，要么平台强，要么大模型强，最好是两者都强，才能在未来 AI 的竞技场上提高胜率。

图14：几个平台的比较

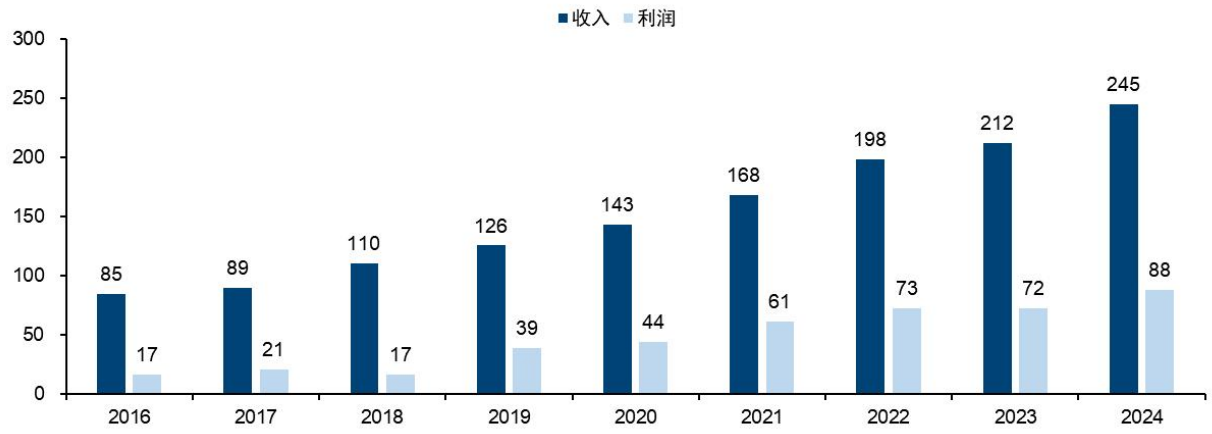
	微软		谷歌	脸书	亚马逊
应用	第三方	Copilot	谷歌应用	脸书应用	第三方
平台	Azure MaaS		Vertex AI	Meta AI	Bedrock
大模型	OpenAI		Gemini	Llama	第三方
云	Azure		谷歌Cloud	Meta	AWS
芯片	英伟达		TPUs	英伟达	英伟达

资料来源：stratechery.com，国信证券经济研究所整理

对于微软来讲，其机会是不言而喻的——纳德拉一直以来的重视，连续多年的投入，OPENAI 的先机，产品线与 AI 结合的经验...目前 OpenAI 已经转成盈利架构，然而这相当于它不再是从前那个研究型公司，而短期估值较高的它，可能将为了盈利，而不会再对隐私保护、AI 发展安全性的那么重视，而是全力实现商业化，未来一旦出隐私、AI 安全等问题，可能对公司的品牌影响与信任度会构成新的压力。

无论如何，2016 年以来，市值翻了 10 倍的微软，证实了纳德拉的眼光和执行力。他没有在硬件上继续同苹果、安卓纠缠，而是把精力放在了云计算和 AI 上，这使得微软在云计算上赢得了新的机遇，同时在 AI 上也占领了先机。此外，他对微软的文化重塑，以及对外合作包容的态度的变化也顺应了时代发展的需要。

图15：2016-2024 财年微软的收入与利润，十亿美元



资料来源：Factset，国信证券经济研究所整理

AMD：轻装出发（fabless）的挑战者

桑德斯是 AMD 的创始人，在他带领下的 30 多年时间里，AMD 成为了英特尔的替代公司，经常是一些客户的第二供应商，在 2000 年科网泡沫时，市值也曾摸到过近 150 亿美元，但桑德斯此时已经 64 岁了，他决定挑选接班人。

鲁伊兹时代（Hector Ruiz，任期 2000-2008 年）：收购 ATI

桑德斯看中了鲁伊兹（Hector Ruiz）——他在德州仪器公司工作了六年，在摩托罗拉公司工作了 22 年，后来升任摩托罗拉半导体产品部门总裁。鲁伊兹在 2000 年加盟 AMD，任 COO，并于 2002 年接替桑德斯担任 CEO，2004 年被任命为董事会主席。

鲁伊兹为 AMD 做了几件重要的工作：

一是在 64 位芯片上取得突破。2003 年 4 月 AMD 早于英特尔发布了 64 位服务器芯片 Opteron，这让 AMD 成功进入高端服务器市场，其服务器市场份额也从 2005 年的 5-7%提升到了 2006 年的 22%;2003 年 9 月,AMD 发布了 64 位 PC 芯片 Athlon 64，其后又推出主打游戏性能的 Athlon 64 FX，两者的性能不输于奔腾 4，甚至某些方面超越了奔腾，且性价比更高。

英特尔此时尚未推出 64 位芯片，它们使用了多种手段与 AMD 竞争：包括宣传 64 位芯片不成熟；与主要的制造商签订排他性协议；提供补贴和激励措施，鼓励 OEM 厂商购买和推广其产品；对 AMD 提起专利侵权诉讼，通过法律手段限制 AMD 的技术发展。

其最后的结果是，AMD 反过来起诉英特尔利用垄断地位不正当竞争，最终虽然在 2009 年，AMD 和英特尔最终达成和解，英特尔同意停止向企业提供回扣并向 AMD 支付 12.5 亿美元的赔偿金，但时隔多年 AMD 一路失去的市场份额却无法找回。

二是收购了 ATI。2006 年 7 月，AMD 宣布以 54 亿美元收购 ATI 公司，其中 42 亿美元为现金，摩根斯坦利为 AMD 提供了 25 亿美元的贷款以完成交易。鲁伊兹考虑的是通过收购 ATI 进入到显卡市场，事实证明这个方向是正确的，但其在收购案中使用了过多的现金，加之在与 ATI 公司的整合遇到文化差异、大客户订单丢失、技术整合时间长等问题，整合进度不及市场预期，以至于公司连续 7 个季度亏损，这也成为鲁伊兹引咎辞职的导火索。

三是巴萨罗那芯片问题。2007 年，AMD 公司推出了 64 位服务器芯片巴塞罗那。但是犯了一个错误（三级缓存缺陷），这会引发服务器的死锁。但当时由于芯片硬件已经没法改动，只能通过软件打补丁的方式解决，而代价是导致 5%到 20%的性能损失，这与此前公司宣传的“比同类英特尔至强双处理器表现出 40%的性能优势”不相符。而此时英特尔已经推出了多款高性能的四核处理器，如 Xeon 系列。相较下来，巴塞罗那芯片的问题影响了 AMD 的品牌形象和市场信心，其服务器市场份额从 2006 年巅峰时期 23%跌至 2008 年的 10-12%之间。

四是推动格罗方德的剥离。晶圆代工是重资产的商业模式。2008 年的 AMD 已经无法在 ATI 与代工双管齐下了。相较下来，公司希望走 fabless 模式。鲁伊兹在任上推动了格罗方德的剥离，2008 年 7 月他辞去了 CEO。2009 年 3 月，AMD 将格罗方德出售给阿布扎比先进技术投资公司（ATIC）和穆巴达拉发展公司（Mubadala），辞职的鲁伊兹任格罗方德的董事长，这为 AMD 的轻资产化以及未来与台积电的合作奠定了基础。

梅耶时代（Derrick Meyer，任期 2008-2011 年）：剥离格罗方德

梅耶是鲁伊兹时期的 COO，鲁伊兹辞职后，他成为了 CEO。梅耶将公司的重点放在个人电脑和数据中心服务器市场上。尽管 2010 年之后，移动端的增长爆发，但是梅耶解释说，移动和消费电子市场的不断增长不会使传统市场萎缩。这个背景主要是因为当时的 AMD 还在亏损中，公司还在急于扭亏，所以无法在 PC 端、移动端大举投入研发。

梅耶在任 3 年的主要贡献有：

1、2009 年完成了前述的对格罗方德的剥离，使得 AMD 有机会选择与台积电合作，公司不需要再背负沉重的代工资产以及投入巨额的世代升级费用，转而专注投入研发；

2、剥离 ATI 公司的手机与电视业务，集中资源和精力在核心的高性能计算和图形市场上；

3、推出三款重要产品，包括 2008 年的 PhenomII 处理器，2009 年的 Athlon II 处理器，以及 Radeon HD 5000 系列显卡。许多专业媒体和评测网站对 Radeon HD 5000 系列给予了高度评价，特别称赞其在 DirectX 11 游戏中的表现和能效比。

2009 年-2010 年，凭借 AMD 的多款产品的发布，AMD 的股票反弹了 5 倍，但是随着欧债危机的到来，加之公司专心耕耘 PC 业务与显卡业务，使得在移动互联网侧并未有多大建树。股价随之也大幅回调，2011 年梅耶辞职，业界认为是董事会认为梅耶在移动互联网端的布局过于谨慎。

平心而论，梅耶上任后止住了 AMD 的失血，让公司在 2009-2011 年连续三年盈利，而且聚焦了最可能快速做强的产品线 PC 与显卡，可以说表现是不俗的。但是随着苹果、安卓手机的大爆发，资本市场认为只有移动端才代表未来，股东们想乘风移动互联网，梅耶则认为而大规模多管齐下投入研发则短期见不到回报，饭要一口一口吃，这种矛盾最后导致他在 2011 年 1 月份辞去了 CEO。

从这个案例可以看出，股东们最大的关心是市值。而在市值成长的路上稍有颠簸，他们的耐心就可能快速消失。

里德时代（Rory Read，任期 2011-2014 年）：面向低功耗与游戏市场转型

里德在 IBM 工作了 23 年，在来 AMD 之前在联想任总裁兼首席运营官。2011 年 8 月里德被任命为总裁兼 CEO，并担任公司董事会成员。

里德接任 CEO 时期，AMD 正面临严峻的市场挑战，包括 PC 市场需求下滑、与英特尔竞争加剧等，且公司 95%的收入都在 PC 市场。他决定减少对传统 PC 市场的依赖，转向更为多元化的市场，特别是移动设备和云计算领域。

里德实施了多项成本节约措施，包括裁员千人，以减少开支并提高公司的财务稳定性，同时在 2012 年发布了推出了 Trinity APU，这是 AMD 首款集成高性能 CPU 和 GPU 的处理器，标志着公司在融合架构上的重要进展。由于收购了 ATI，因此 APU 相当于将 AMD 的独特定位（CPU+GPU）与价值凸显出来。

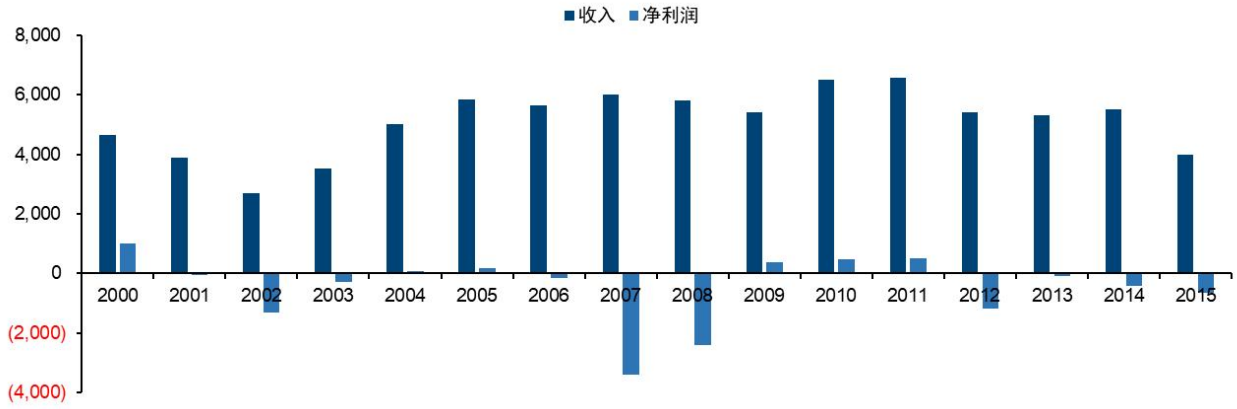
2013 年，公司发布了 Kabini 和 Temash 处理器，Kabini 主要面向笔记本电脑和平板电脑市场，主打低功耗，低成本；后者面向游戏电脑和工作站市场。同年，公

司推出了基于 Jaguar 核心的 Opteron 处理器，这是 AMD 首次推出专门针对低功耗服务器市场的处理器。

此外，里德开始将半定制业务提升为公司的重要战略方向之一。2012 年，公司赢得了 PlayStation 4 和 Xbox One 的合同，两者采用了基于 Jaguar 架构的 8 核 CPU 和 GCN 架构的 GPU。由于 PS4 和 Xbox One 的市场表现出色，为 AMD 带来了稳定的收入来源，这两款游戏机的成功也进一步巩固了 AMD 在游戏机市场的地位。

2013 年 9 月，AMD 发布的基于 Hawaii 芯片的 GPU 产品线也表现不俗（对标英伟达 GeForce GTX 980 和 GTX 970），在 2014 年占据了全球 GPU 市场的约 30%的份额。

图16：2000-2015 年，AMD 收入及利润（百万美元）



资料来源：Factset，国信证券经济研究所整理

里德还有一个重要贡献，就是 2012 年他将苏姿丰带入 AMD。2014 年，里德离职，苏姿丰接任 CEO。虽然三年时间 AMD 并未大踏步前进，但是它逐渐在产品线上找到了新的方向，特别是低功耗技术与新兴市场的拓展，为 AMD 后续在移动设备和数据中心市场的成功奠定了基础。

苏姿丰时代（Lisa Su，任期 2014-今）：再次腾飞

终于，AMD 迎来了苏姿丰时代。

苏姿丰出生于台湾，儿时移居美国。在麻省理工学院获得三个学位后，她曾在德州仪器、IBM 和飞思卡尔半导体公司任职。在担任 IBM 半导体研发中心副总裁期间，她因开发绝缘体上硅半导体制造技术和更高效的半导体芯片而闻名。因此，苏姿丰是典型的技术出身的管理层。

苏姿丰于 2012 年加入 AMD，担任过 AMD 全球业务高级副总裁和 COO，并于 2014 年 10 月被任命为 CEO。2023 年，苏姿丰在《福布斯》“全球 100 位最具影响力女性”榜单中排名第 49 位。在《财富》杂志 2023 年最具影响力女性榜单中排名第 12 位。《时代》杂志将她列入 2024 年“人工智能领域最具影响力的 100 人”榜单。

图17：2015-2023 年，AMD 收入及利润（百万美元）

资料来源：Factset，国信证券经济研究所整理

从苏姿丰接任 AMD 开始到 2024 年，AMD 的股价最大上涨了 80 倍，目前稳定在 50 倍左右。在她领导下的 AMD 再次腾飞，甚至缔造了其诸多前任都未能达到的历史高度，一个最重要的变化是，AMD 从第二梯队的“跟随者”，逐渐跃入第一梯队。在这个跃升的过程中，可以分成三个阶段：

1、“ZEN”架构（2015-2018 年）

AMD 的股价在 2015 年创下新低，原因包括英伟达与英特尔的竞争压力、公司的库存问题也凸显。刚上任的苏姿丰提出，公司应该专注于利润率更高、增长机会更高的市场。**当时 AMD 规划了三大增长市场：游戏、沉浸式平台和数据中心市场。**同时提出 AMD 应避免进入利润率低的市场，如移动（智能手机/平板电脑）或物联网领域，因为尽管这些领域虽然增长迅速，但竞争对手已经太多，这些竞争对手要么拥有成本控制能力（联发科），要么拥有资金雄厚（英特尔），可以将利润率压低到 AMD 无法维持的水平。同样，AMD 正在努力减少其在低端 PC 市场的份额，因为该市场的利润率也很低，再加上前景不佳，AMD 因该市场需求大幅下降而遭受重创。其中，沉浸式平台主要指的是 AR/VR，但其后这个方向慢慢在发展过程中被搁置。

苏姿丰的大方向非常正确。前三任 CEO 总体上从都是“腾笼换鸟”，没有人敢从提升利润率的角度来思考：因为高利润率对一个科技公司至关重要，因为高利润率很大程度上代表了技术领先，但当时的 AMD 还没有底气能够拿出最先进的技术以打动市场。

这就不得不提到 Zen 架构。Zen 架构于 2012 年开始投入研发，到 2017 年正式面世，期间经历了 5 年的时间。它是一种全新架构，与以往架构的不同是：它具有更高的性能、更低的功耗和更好的可扩展性。AMD 在全球范围内调集了数千名工程师和技术专家，参与 Zen 架构的研发工作，这些团队分布在北美、中国、印度等地，涵盖了 CPU 设计、GPU 设计、软件开发等多个领域。

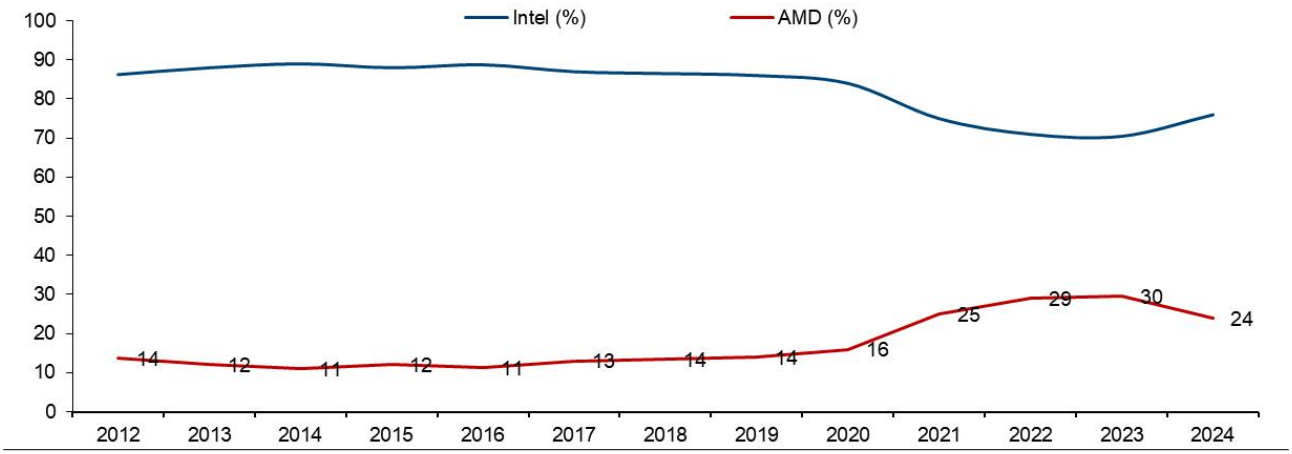
我们在此前的报告曾提及，英特尔对 AMD 的市场主导地位随着“酷睿”品牌的推出，以及“滴答（Tick-tock）”发布策略的成功推出而不断增强，该模式最著名的是每年在新的 CPU 架构和新的制造节点之间交替。英特尔遵循该发布节奏近十年，从酷睿于 2006 年首次推出 65 nm 的 Conroe 架构开始，一直持续到 14 nm 的 Broadwell 架构，从 Broadwell 开始，其在 2014 年的发布计划推迟了一年到 2015 年的 Q3，这宣告了“滴答”模式的终结。这件事对 AMD 来说至关重要，因为英特尔无法进一步维持“滴答”模式，对于 AMD 的锐龙 CPU 以及整个 Zen 架构的成功

提供了新的市场机遇。

在经历了 2015-2016 相对低谷的两年后，Zen 架构于 2017 年 2 月首次推出，首发于第一代锐龙（Ryzen）CPU。该架构用于也用于锐龙（台式机和移动设备）、锐龙 Threadripper（工作站和高端台式机）和 Epyc（服务器）。

基于 Zen 架构的 Ryzen 处理器正式推出，性能大幅提升，例如 Ryzen 7 1800X 的性能略优于 Core i7-6900K，但价格只是酷睿的一半，且 Ryzen 在能效方面也有优势，因此 AMD 从 2017 年开始，开始重新夺回 CPU 市场份额。

图18：2012-2024 全球 CPU 市场份额



资料来源：pcviewed.com，国信证券经济研究所整理

尤其是基于 Zen 的服务器芯片 EPYC 发布后，AMD 获得了惠普、戴尔、SuperMicro、赛灵思、VMWare、Red Hat、微软等多个客户的支持，第二代 EPYC 定位为数据中心和云，也迅速赢得了业界的广泛好评，谷歌也成为客户。2014 年 Q4 苏姿丰上任时，AMD 在服务器芯片份额仅为 1.1%，2018 年则升至 3.7%。

此时，苏姿丰的战略也逐渐清晰且自信，她认为高性能计算和数据中心是公司更为清晰的未来战略。

下表是 AMD 的 Zen 架构主要 CPU 与英特尔的比较，可以看出，在 2017 年发布的第一年，两者的价格几乎相差一半，而随后由于 Zen 的性能和口碑开始被市场认同，英特尔大幅降价，两者的价格差距快速缩小，甚至在 2020 年，AMD 售价还略超过了英特尔。

表3：2017 年以来，AMD 与英特尔 CPU 售价比较

Zen 世代	Zen	Zen+	Zen 2	Zen 3	Zen 3+	Zen 4	Zen 5
年份	2017	2018	2019	2020	2022	2023	2024
AMD	Ryzen 7 1800X	Ryzen 7 2700X	Ryzen 9 3900X	Ryzen 9 5900X	Ryzen 7 5800X3D	Ryzen 9 7950X	AMD Ryzen 9 9950X
制程工艺	14nm 格罗方德	12nm 格罗方德	7nm 台积电	7nm 台积电	7nm 台积电	5nm 台积电	4nm 台积电
价格，美元	499	329	499	549	449	549	649
INTEL	Core i7-6900K	Core i7-8700K	Core i9-9900K	Core i9-10900K	Core i9-12900K	Core i9-13900K	Core Ultra 9 185H
制程工艺	14nm	14nm++	14nm++	14nm++	Intel 7	Intel 7	Intel 4
价格，美元	1089	359	499	488	589	589	640
价格差（英特尔-AMD）	590	30	0	-61	140	40	41

资料来源：英特尔，AMD，国信证券经济研究所整理

2、台积电代工（2019-2021 年）

终于在剥离了格罗方德之后，从 Zen 2 开始，AMD 开始选择台积电代工 7nm 制程的 CPU，这对 AMD 的产品又有了巨大的性能提升。由于英特尔的大部分高端处理器，如 Core i7、i9 以及服务器用的 Xeon 处理器，仍然主要由英特尔自己制造。这些处理器采用英特尔的制程工艺，如 Intel 4（原 7 纳米改进）、Intel 7（原 10 纳米改进）等，因此英特尔在制程上的瓶颈成为影响公司芯片性能表现的“短木板”。

从 2019 年开始，AMD 将代工从格罗方德转向台积电之后，可以看出，Zen 的制程工艺始终领先英特尔 1-2 代。Zen 2、Zen 3、Zen 3+采用了 7nm 制程，而当时的 Core i9-9900k、10900k 制程还停留在 14nm 的改进版，12900k 则是采用 10nm 工艺（Intel 7）；Zen 4 采用了 5nm 制程，Core i9-13900k 则采用 10nm 工艺（Intel 7），直到 2013 年 12 月份，英特尔发布的 Core Ultra 系列芯片才升级到了 7nm 工艺（Intel 4）。

2019 年，Zen 2 架构的锐龙 3000 发布，服务器芯片霄龙（EPYC）第二代发布；2020 年锐龙 4000 发布，它针对笔记本市场，同时服务器芯片采用 Zen 3 的霄龙米兰（EPYC Milan）第三代发布。

由于制程上的领先，AMD 在高性能、低功耗以及集成化 APU（CPU+显卡）的优势逐渐显露，2019-2021 其市场份额也大幅增长。CPU 份额从 2019 年的 14%提升至 2021 年的 25%，收入从 2019 年的 67 亿美元大幅提升至 2021 年的 164 亿美元，利润从 2019 年的 3.4 亿美元，提升至 2021 年的 31.6 亿美元，股价从 2018 年的低点计算，三年翻了 10 倍左右。

如果说，2015-2018 年，AMD 股价上涨了 5 倍主要是归功于 Zen 架构的诞生，那么 2019-2021 年，AMD 的股价上涨 10 倍则是归于 Zen 架构与台积电制程结合下迸发出的巨大优势，在这三年中，公司大幅修复了现金流量表，同时员工人数也从不到 10000 人恢复到了 15000 人。

表4：AMD 的 Zen 架构家族

架构	芯片	发布时间	面向市场
Zen 1	Ryzen 1000 系列（Summit Ridge）	2017	桌面
	Threadripper 1000 系列（Whitehaven）	2017	桌面
	Ryzen 2000 系列（Raven Ridge）	2018	桌面、移动
	Ryzen 3000 系列（Dalí ）	2019-2020	移动
	V1000 系列（Great Horned Owl）	2018	嵌入式
	R1000 系列（Banded Kestrel）	2019-2020	嵌入式
	Epyc 7001 系列（Naples）	2017	服务器
Zen+	Ryzen 2000 系列（Pinnacle Ridge）	2018	桌面
	Threadripper 2000 系列（Colfax）	2018	桌面
	Ryzen 3000 系列（Picasso）	2018	桌面、移动
	R2000 系列（River Hawk）	2022	嵌入式
Zen 2	Ryzen 3000 系列（Matisse）	2019	桌面
	Threadripper 3000 系列（Castle Peak）	2020	桌面
	Ryzen 4000 系列（Renoir）	2021-2022	桌面、移动
	Ryzen 5000 系列（Lucienne ）	2021	移动
	Ryzen 7000 系列（Mendocino ）	2022	移动
	V2000 系列（Grey Hawk）	2020	嵌入式
	Epyc 7002 系列（Rome）	2019	服务器
Zen 3	Ryzen 5000 系列（Vermeer）	2021-2024	桌面
	Ryzen 5000 系列（Cezanne）	2021-2024	桌面、移动
	Threadripper 5000 系列（Chagall）	2022	桌面
	Ryzen 7000 系列（Barcelo-R）	2023	移动
	V3000 系列	2022	嵌入式
	Epyc 7003 系列（Milan）	2021	服务器
Zen 3+	Ryzen 6000 系列（Rembrandt ）	2022	移动
	Ryzen 7000 系列（Rembrandt-R ）	2023	移动
Zen 4	Ryzen 7000 系列（Raphael）	2022-2023	桌面
	Threadripper 7000 系列（Storm Peak）	2023	桌面
	Ryzen 8000 系列（Phoenix）	2024	桌面
	Ryzen 7000 系列（Phoenix）	2023	移动
	Ryzen 7000 系列（Dragon Range）	2023	移动
	Ryzen 8000 系列（Hawk Point）	2023	移动
	Ryzen Z1 系列	2023	游戏机
	Ryzen Embedded 7000 系列	2023	嵌入式
	Epyc 9004 系列（Genoa）	2022	服务器
	Epyc 9004 系列（Bergamo）	2023	服务器
	Epyc 8004 系列（Siena）	2023	服务器
Zen 5	Ryzen 9000 系列（Granite Ridge）	2024	桌面
	Ryzen AI 300 系列（Strix Point）	2024	移动
	Epyc 9005 系列（Turin）	2024	服务器

资料来源：AMD，国信证券经济研究所

3、发力人工智能（2022-2024 年）

EPYC 服务器系列的成功，让 AMD 更加坚信，只有通过技术创新来提升产品的竞争力，让客户之间有口碑认同，才能不断地提升市场份额，而再投入更多的研发以促进创新。或者说，从苏姿丰开始，AMD 慢慢找到了这种可以清晰的、持续的、自我强化的战略，他们将技术创新的方向瞄准到高性能运算（High Performance），那么就不能不提到 AI 芯片市场的巨大机会。

由于竞争对手英伟达早在 2006 年就推出了 CUDA，在十几年间，英伟达 GPU 早已构筑出宽广的护城河。

AMD Instinct 是 AMD 的数据中心 GPU 品牌。它在 2016 年取代了 AMD 的 FirePro S 品牌。与消费级 Radeon 相比，Instinct 产品线面向的是加速深度学习、人工神

经网络和高性能计算/GPGPU 应用。

从 2020 年开始，AMD 的 CDNA 架构（Compute Data Node Architecture）替代了 GCN（Graphics Core Next）架构，CDNA 架构拥有更高的计算密度和能效、增强的内存子系统，支持 HBM3，优化了优化的数据路径和互连，具有高级计算特性，增加了硬件加速功能，也得到了 AMD 的 ROCm（Radeon Open Compute）开源平台的全面支持，提供了丰富的开发工具和库，方便开发者进行高性能计算和 AI 应用的开发，同时在可扩展性和模块化设计也有考虑。这些优势使得 CDNA 架构能够更好地满足现代计算和 AI 应用的需求。

表5：AMD 的 AI 芯片 Instinct 系列

芯片	发布时间	架构	制程	计算单元	内存	FP32 算力	功耗 W
MI6	2016	GCN 4	14nm	36	16GB GDDR5	5.7 TFLOPS	150
MI8	2016	GCN 3	28nm	64	4GB HBM	8.2 TFLOPS	175
MI25	2016	GCN 5	14nm	64	16GB HBM 2	12.3 TFLOPS	300
MI50	2018	GCN 5	7nm	60	16GB HBM 2	13.3 TFLOPS	300
MI60	2018	GCN 5	7nm	64	32GB HBM 2	14.7 TFLOPS	300
MI100	2020	CDNA	7nm	120	32GB HBM 2	23.1 TFLOPS	300
MI210	2022	CDNA2	6nm	104	64GB HBM2e	22.6 TFLOPS	300
MI250	2021	CDNA2	6nm	208	128 GB HBM2e	45.3 TFLOPS	560
MI250X	2021	CDNA2	6nm	220	128 GB HBM2e	47.9 TFLOPS	560
MI300A	2023	CDNA3	5-6nm	228	128 GB HBM3	122.6 TFLOPS	550
MI300X	2023	CDNA3	5-6nm	304	192 GB HBM3	122.6 TFLOPS	550
MI325X	2024	CDNA3	4-5nm	304	288 GB HBM3	163.4 TFLOPS	750
MI350X	2025e	CDNA4	3nm		288 GB HBM3e		
MI400X	2026e						

资料来源：AMD，国信证券经济研究所整理

由于在 2022 年 Chat GPT 3.5 的出现，全面引爆了 AI 芯片市场，使得积累多年的英伟达收入增长一骑绝尘。AMD 也在 2023 年发布了 MI 300X 芯片，并在 2024 年取得了非常丰厚的回报。2024 年 Q2，公司数据增速 28.34 亿美元，同比增长 115%。以 AI 芯片为例，2023 年 10 月公司预期 2024 年 AI 芯片收入为 20 亿美元，2024 年 1 月预期提升至 35 亿美元，2024 年 7 月又提升至 45 亿美元，2024 年三季报后又近一步提升至 50 亿美元。

表6：AMD 收入结构说明

业务分项	说明	子项目
1. 数据中心	数据中心、云计算和企业服务器相关的产品和服务	服务器 CPU（EPYC 系列） 数据中心 GPU（Radeon Instinct 系列） 存储产品（SSD 等） 数据中心相关软件和服务
2. 游戏	游戏相关的产品和服务，包括图形卡、游戏主机和游戏软件。游戏主机（半定制 GPU 游戏主机，如 PlayStation、Xbox 等）	图形卡（Radeon RX 系列、Radeon Pro 系列） 游戏相关软件和服务
3. 嵌入式	嵌入式系统相关的产品和服务，包括 CPU、GPU、微控制器和其他嵌入式产品。	嵌入式 CPU（Ryzen 嵌入式系列、EPYC 嵌入式系列） 嵌入式 GPU（Radeon E 系列等） 微控制器（MCU）和其他嵌入式产品 嵌入式相关软件和服务
4. 客户端	AMD 与客户端计算相关的产品和服务，包括台式机和笔记本电脑的 CPU，以及与客户端相关的软件和服务。	桌面 CPU（Ryzen 系列、Athlon 系列） 笔记本 CPU（Ryzen 系列、Athlon 系列） 客户端相关软件和服务

资料来源：AMD，国信证券经济研究所整理

AMD 在 2024 年预告了 MI350 系列，以对标英伟达 Blackwell 系列，其性能提升高达惊人的 3500%，预期将在 2025 年下半年发布。苏姿丰估计，AI 芯片市场规模将

以超过 60%的 CAGR 增长，并于 2028 年达到 5000 亿美元。

目前，AMD 在服务器 CPU 市场份额 EPYC 系列已经超过 25%，但 AI 芯片市场份额仅为 4-7%左右（英伟达 90%+），倘若在 2028 年，其份额可以像 EPYC 系列那样达到 20%左右，那么光 AI 芯片就有 1000 亿美元收入。但截止 2024 年 Q3，AMD 在 AI 芯片市场的份额仅为 5%，因此市场还不敢过度乐观认为这一目标能够轻易实现。过去 5 年，市场给 AMD 的平均市盈率不够稳定，市销率大约在 9 倍左右。

英伟达的成功不仅仅在于其芯片，更在于其软件栈 CUDA 的成功，CUDA 已成为人工智能开发者的标准语言。CUDA 允许软件开发者使用英伟达 GPU 加速并行通用计算，不兼容 AMD 以及英特尔。面对领先的英伟达以及 CUDA，AMD 的芯片与 ROCm 诞生于 2016 年，整整晚了 10 年，ROCm 是免费、自由和开源的软件。近一年来，AMD 打通了所有主要 AI 开发平台，获取了 PyTorch 的零日更新支持及 Triton 的 AMD 硬件兼容。

此外，在收购的路上，最近几年 AMD 也是大动作不断。比较重要的收购包括：

1、2022 年 2 月完成对赛灵思（Xilinx）的收购，最终价值为 498 亿美元。该并购强化了 AMD 在高性能计算领域的地位，并为其带来了在数据中心、边缘计算以及其他高性能计算应用领域的更多机会；

2、2022 年 4 月宣布 19 亿美元收购 Pensando，它的分布式服务平台包括一个高性能的、完全可编程的分组处理器和全面的软件栈，可以加速云计算、企业和边缘应用的网络、安全、存储和其他服务。客户包括高盛、IBM 云、微软 Azure 和甲骨文云；

3、2024 年 7 月宣布将以 6.65 亿美元全现金收购 Silo AI。Silo AI 是欧洲最大的私人人工智能（AI）实验室，业务遍及欧洲和北美。此次收购代表了 AMD 基于开放标准并与全球 AI 生态系统建立强有力的合作伙伴关系，并提供端到端 AI 解决方案的战略又迈出了重要一步；

4、2024 年 8 月，宣布以 49 亿美元收购 ZT Systems。此次收购将增强 AMD 在数据中心 AI 系统和客户支持方面的能力，同时公司希望布局人工智能软件堆栈，AMD 将把各种元素结合在一起，提供一个真正的人工智能解决方案路线图。

2016 年之后，“tick-tock”止步不前的英特尔暗示了芯片制造领域重资产模式的受阻，也使得其想通过收购 Altera 并未达到预想的效果。而英伟达与剥离了格罗方德的 AMD 则是凭借着轻资产模式发展迅猛，其最高市值已经超过了 3600 亿美元，而宿敌英特尔则跌破了 1000 亿美元。

AMD 得以在稳固 CPU 市场之后，开始像当年投入 Zen 架构一样，一点点缩短在 AI 市场上与英伟达的差距。考虑到未来三五年，大模型将走向十万亿、百万亿参数级别，其算力潜力需求巨大而客户会不得不更多考虑性价比因素，因此从这个角度说，居第二的 AMD 只要不掉队，凭借性价比的优势，如果在客户认同上能明显缩小与英伟达的差距，则会获得更大的空间。

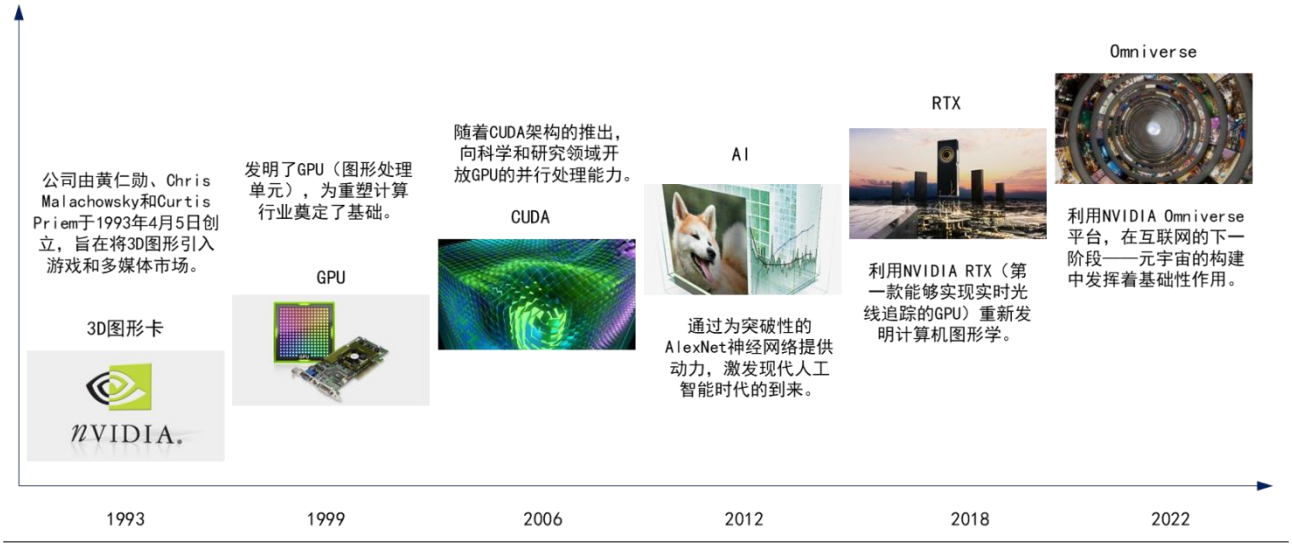
英伟达：摩尔定律的延续者

CUDA 标志着“指数型思维”的思想延续（2006 年）

与其他案例不同，英伟达的 1993 年成立，创始人黄仁勋一直担任公司的 CEO。到 2024 年，他已经执掌公司 31 年。我们知道，乔布斯领导下的苹果，盖茨领导下的微软，格鲁夫/摩尔博士领导下的英特尔，佩奇领导下的谷歌，戴尔领导下的戴尔电脑...以及我们在《移动互联网案例篇》中举到的诸多例子都表示，初代创始人熟悉行业的来龙去脉，同时他们对技术或者产品敏锐，也非常了解自己公司的优点与不足。因此在他们领导下的公司，大多数案例都是成功的。而等到二代、三代 CEO，有的好，有的一般，但总体来说遇到杰出 CEO 的概率不高，他们很难超过创始 CEO 的水平。

黄仁勋 1963 年出生在台南，9 岁赴美学习，1984 年于俄勒冈州立大学获取电机工程学士学位，1992 年于斯坦福大学获取电子工程学硕士学位。1983 年黄仁勋在 AMD 担任微处理器硬件工程师，1985 年至 1993 年，他在 LSI Logic 担任核心硬件设计总监。并于 1993 年与 Chris Malachowsky（目前依然在公司任职）和 Curtis Priem（2003 年从英伟达退休）共同创办了英伟达，并任 CEO。

图19：英伟达的发展简史



资料来源：英伟达，国信证券经济研究所整理

我们在《科技周期探索之三：1974-1987 年：个人电脑时代的到来》中的总结曾提及：千万不要忽视“指数型思维的人或者公司”。当时提到的案例是有着摩尔博士掌舵的英特尔，在他的任期中，英特尔的股票翻了上百倍，而他退休之后英特尔就很难找回当初的状态。

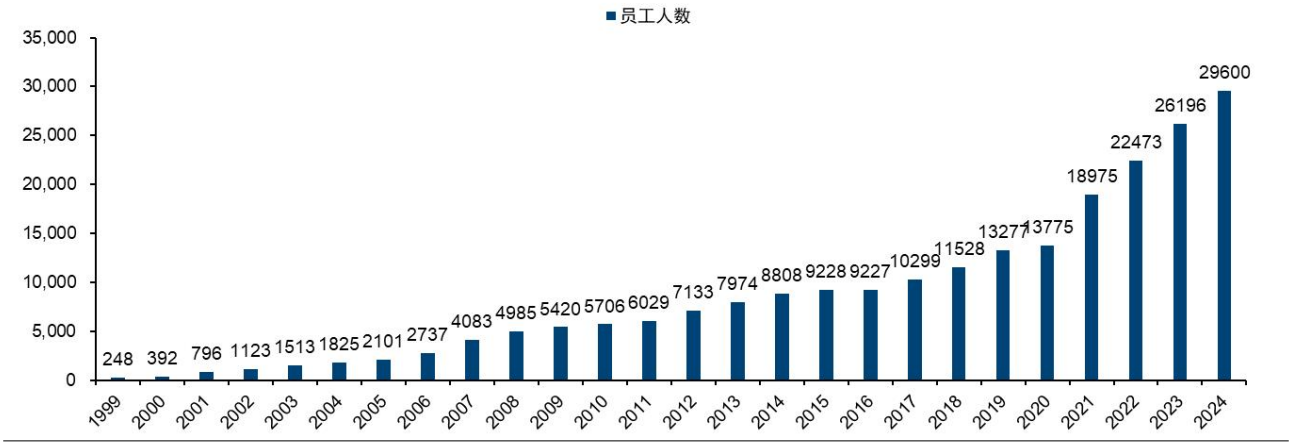
黄仁勋和英伟达，就是又一个具备“指数型思维的人或者公司”。甚至说就是因为英伟达的存在，GPU 接替了 CPU，才让摩尔定律曲线得以延伸到现在也不为过。

从 2006 年 CUDA 推出之后，英伟达就不再是一家显卡公司，它成为了通用 GPU 的供应商。由于当时的 CUDA 思想过早，导致了业界好多年都没有看懂，也没有试着模仿——直到 2016 年 AMD 才开发了自己的 ROCm，2019 年英特尔的 oneAPI 正式版才推出。因此说，CUDA 的推出，以及通用 GPU 架构的出现，是英伟达的重要壮举。

在成绩面前，人们都容易归功于 CEO 的战略；但 2008 年金融危机时期，尽管面临经济危机，CUDA 的推出被视为“巨大的豪赌”，当时公司的股价从高点跌去了 80%，市值仅存 40 亿美元，但英伟达依然坚持技术创新，保住了 CUDA 这颗冉冉升起的新星。

经历科网泡沫、金融危机、欧债危机、新冠疫情，英伟达从来没有“净裁员”，员工人数一直在增加，这在纳斯达克是极为罕见的——充分说明管理层高度的前瞻性和面对创新不确定性下的定力。

图20：英伟达的员工人数



资料来源：英伟达，国信证券经济研究所整理

从 2006 年的 Tesla 架构到 2010 年的 Fermi 架构，英伟达与 AI 的关系还没有很大，主要是在摸索通用 GPU 的一些基础性能，包括统一着色器、增加 CUDA 核心，以及支持 DirectX 11 的架构等等。

不冷不热的移动互联网尝试（2010-2015 年）

到了 2011 年，橡树岭国家实验室在建造超级计算机 Titan（泰坦）时，大量采购了英伟达的 GPU，泰坦成为世界上第一台使用通用 GPU 的超级计算机。整台泰坦共计 18688 颗 CPU 和相同数量的 GPU，在 2012 年 11 月的测试中获取 17.59petaFLOPS 的成绩，直到 2013 年 6 月在 Top500 位列第一的排名被中国的天河二号取代。泰坦的成功，使得英伟达成为 GPU 行业的一张名片。

表7：英伟达通用 GPU 的架构

架构名称	发布时间	制程工艺	代表型号	功率范围	说明
Tesla	2006 年	90nm	GeForce 8800 GTX	175W	引入统一着色器架构
Tesla	2008 年	65nm/55nm	GeForce GTX 280	236W	增强了 CUDA 核心
Fermi	2010 年	40nm	GeForce GTX 480	250W	第一个支持 DirectX 11 的架构
Kepler	2012 年	28nm	GeForce GTX 680	195W	提升了能效比
Maxwell	2014 年	28nm	GeForce GTX 980 Ti	250W	优化内存带宽
Pascal	2016 年	16nm	GeForce GTX 1080 Ti	250W	高性能与能效
Volta	2017 年	12nm	Tesla V100	300W	面向数据中心，首次引入 Tensor Core，专注于深度学习和 AI 应用
Turing	2018 年	12nm	GeForce RTX 2080 Ti	250W	引入了 RT Core，支持实时光线追踪
Ampere	2020 年	8nm	GeForce RTX 3090	350W	支持 RTX IO
Hopper	2022 年	4nm	H100	700W	专为 AI 和数据中心设计
Blackwell	2024 年 3 月	3nm	GB200	1000W	专为 AI 和 HPC 设计，支持 HBM3E 显存，具有高带宽和低功耗特性。

资料来源：英伟达，国信证券经济研究所整理

2012 年，AlexNet 在 ImageNet 挑战赛中取得了突破性成绩，这一成就的背后是英伟达 GPU 的支持。AlexNet 使用的是英伟达 GTX 580 GPU，基于 Fermi 架构。

Kepler 架构是英伟达在 2012 年推出的 GPU 架构，相较于前一代的 Fermi 架构，它在多个方面进行了显著改进，提供了更高的能效和更强的计算能力，下图可见，基于 Kepler 架构的 GeForce GTX 680，其 CUDA 核心数高达 1536 个，是 GTX 580 的 3 倍。

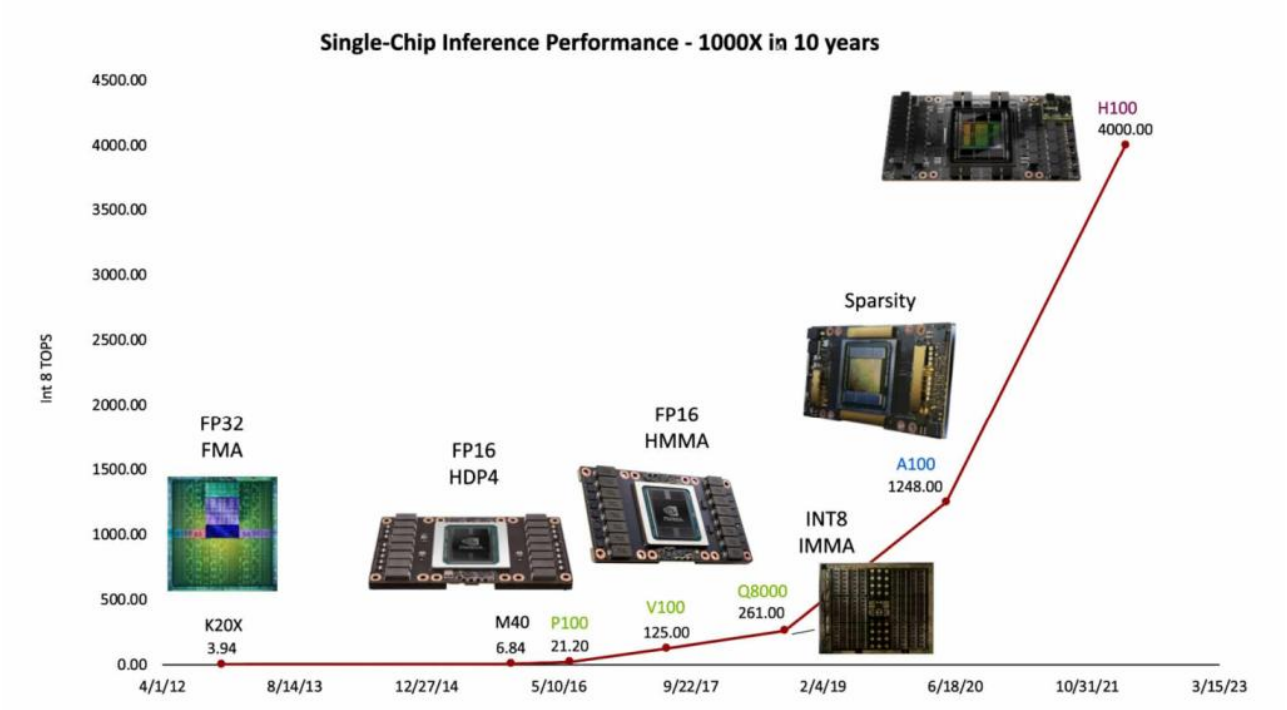
表8：部分英伟达 GPU 的 CUDA 核心数

GPU 型号	发布时间	制程工艺（纳米）	CUDA 核心数
GeForce 8800 GTX	2006 年 11 月	90	不适用
GeForce GTX 280	2008 年 6 月	65	240
GeForce GTX 480	2010 年 4 月	40	480
GeForce GTX 580	2010 年 11 月	40	512
GeForce GTX 680	2012 年 3 月	28	1536
GeForce GTX 780	2013 年 5 月	28	2304
GeForce GTX 980	2014 年 9 月	28	2048
GeForce GTX 1080	2016 年 5 月	16	2560
GeForce RTX 2080 Ti	2018 年 9 月	12	4352
GeForce RTX 3090	2020 年 9 月	8	10496
H100	2022 年 3 月	4	14592

资料来源：英伟达，国信证券经济研究所整理

从 2012 年开始，英伟达开始统计 AI 推理速度的快速进化。

图21：英伟达 GPU 在 10 年的时间里，AI 推理速度提升了 1000 倍



资料来源：英伟达，国信证券经济研究所整理

基于 Kepler 架构的 K20X（从这里开始，芯片的首字母就是架构名称的首字母）的 CUDA 核心数达到了 2688 个。2015 年，公司发布了 Maxwell 架构，它的主要变化是优化了内存带宽，与 K20X 一样，它们同属于 28nm 制程。

在这一段时间里，英伟达从收入、利润上并未有较大的变化。当时的风口在移动互联网侧，公司的确在 2010 年发布了自己的移动互联网芯片 Tegra，该系列处理器是针对移动设备设计的系统级芯片（SoC），应用在智能手机、平板电脑、汽车信息娱乐系统和其他移动设备上，但总体上并没有实现预期的快速增长，原因是：

1、竞争对手强大：高通、三星、苹果等公司在移动处理器市场上拥有强大的技术和市场份额。高通的 Snapdragon 系列处理器在性能、功耗管理和生态系统支持方面表现优秀，吸引了大量 OEM 厂商；

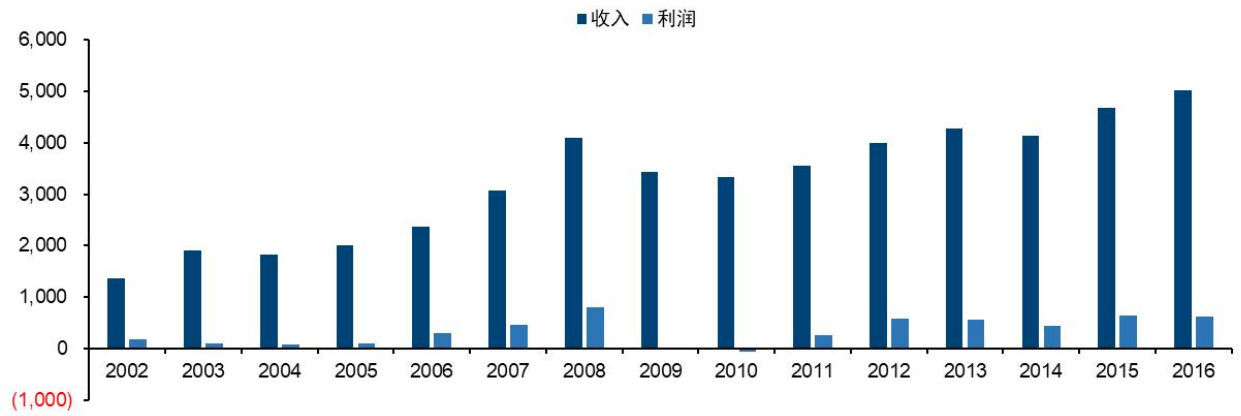
2、缺乏 CPU 的积累：虽然 Tegra 处理器在图形处理能力方面表现出色，但在 CPU 性能上与竞争对手相比存在一定差距，特别是在多任务处理和复杂应用方面。或者说，英伟达是显卡出身的企业，它的优势在高性能运算市场，而此时的平板、手机追求的不是高功耗下的性能，而是在低耗电下的性能，因为随着手机与平板屏幕越来越大，而尺寸越做越薄，手机企业没有余地再将过多资源分配给图形处理了；

3、缺乏生态的支持：同样，摩托罗拉、HTC 和 LG 这些企业不会一下子切换到英伟达，而缺乏长期稳定的 OEM 合作伙伴关系影响了 Tegra 的市场推广；因此说，移动互联网浪潮下的红利不可能看到了才去争取，而像乔布斯那样，从 iPod 时代就已经开始默默地努力，而在 10 年之后大放异彩。

尽管 Tegra 在传统移动设备市场上表现平平，但在汽车信息娱乐系统和自动驾驶领域取得了显著成功。例如，特斯拉和多家高端汽车品牌采用了 Tegra 芯片，因为在汽车的体积下，耗电并不是问题，而性能是企业更关心的，这与数据中心的场景是很相似的，这里是英伟达的强项。

此外，2015 年英伟达发布了嵌入式芯片 Jetson 系列（Maxwell 架构），它针对边缘计算和嵌入式系统设计的模块化计算机，应用于机器人、无人机、智能摄像头等设备。Jetson 系列也算是成功的，因为它为英伟达的边缘运算提供了广阔的市场空间，包括机器人、无人机、工业自动化、智能交通、医疗健康、零售物流、智慧城市、农业等领域，比如亚马逊的配送机器人 Scout 使用了 Jetson Xavier NX 模块，博世的智能工厂解决方案中使用了 Jetson 系列模块。这也将是英伟达在未来 AI 行业落地的一个有效的抓手。

图22：2002-2016 财年（移动互联网时代）英伟达的收入与利润，百万美元



资料来源：Factset（英伟达财报在 1 月发布，因此财报年-1 对应的是自然年，如 2016 年报对应的 2015 年），国信证券经济研究所整理

云计算潮流中崭露头角（2016-2019 年）

到了 2016 年，正值谷歌的 Alpha Go 横空出世，机器学习也成为席卷 AI 界的新潮流。英伟达重要的 Pascal（帕斯卡）架构的 P100 芯片诞生了，它的制程达到了 16nm，CUDA 核心数达到了 3584，其在 INT 8 的算力达到了 21.2TOPS，是 M40 的 3 倍以上。

表9：英伟达 AI 推理部分芯片的发行时间

芯片型号	发行时间	架构	制程工艺	CUDA 核心	功率（W）	发行价格（美元）
K20X	2012 年	Kepler	28nm	2688	235	3199
M40	2015 年	Maxwell	28nm	3072	250	4899
P100	2016 年	Pascal	16nm	3584	250 - 300	5600
V100	2017 年	Volta	12nm	5120	300 - 350	6999
Q8000	2018 年	Tesla	55nm	4608	236	12999
A100	2020 年	Ampere	7nm	6912	250 - 400	19900
H100	2022 年	Hopper	4nm	14592	700	25,000 - 30,000
H200	2023 年	Hopper	4nm	NA	700	40000

资料来源：英伟达，国信证券经济研究所整理

2016 年 4 月，英伟达发布了 DGX 服务器。其中 DGX-1 服务器配备 8 个基于 Pascal 或 Volta 子卡的 GPU，总共 128GB HBM2 内存，通过 NVLink 网状网络连接。与之前的架构相比，Pascal 架构在深度学习任务上提供了 10 倍以上的性能提升，极大地加速了训练过程。DGX-1 预装了优化的深度学习软件，包括英伟达 DIGITS 和 cuDNN，使得研究人员和能够快速而轻松地训练深度神经网络。

图23：英伟达 DGX-1 服务器

资料来源：英伟达，国信证券经济研究所整理

图24：包含了 5 个 DGX-1 的超级计算机的机架

资料来源：维基百科，国信证券经济研究所整理

值得注意的是，NVLink 在 2016 年发布的 DGX 第一次被使用，它是 Nvidia 开发的有线串行多通道近距离通信链路。传统个人电脑的 PCI 接口是串行的，而 NVLink 使用网状网络，对并行运算支持得更好。

由于 PCI 是由 PCI-SIG 联盟（外围组件互连小组，一个电子行业联盟）维护，它沿袭的是 PC 个人电脑主线的标准，无论在组网方式上，还是在传输速率上，都达不到日益增长的机器学习的需要，因此 NVLink 不断迭代，其传输速率已经由 2016 年的 20 Gbits/s 增长到 2024 年的 100 Gbits/s，规划中的 Blackwell 架构则是支持 200 Gbits/s。

表10：NVLink 不同标准的比较

互联标准	传输速率	单向单通道载荷率	实现架构
PCIe 3.x	8 GT/s	约 1 GB/s	Pascal，Volta，Turing
PCIe 4.0	16 GT/s	约 2 GB/s	Volta on Xavier，Ampere
PCIe 5.0	32 GT/s	约 4 GB/s	Hopper
PCIe 6.0	64 GT/s	约 8 GB/s	Blackwell
NVLink 1.0	20 Gbit/s	约 2.5 GB/s	Pascal，
NVLink 2.0	25 Gbit/s	约 3.125 GB/s	Volta，NVSwitch for Volta
NVLink 3.0	50 Gbit/s	约 6.25 GB/s	Ampere，NVSwitch for Ampere
NVLink 4.0	100 Gbit/s	约 12.5 GB/s	Hopper，Nvidia Grace Datacenter/Server CPU，NVSwitch for Hopper
NVLink 5.0	200 Gbit/s	约 25 GB/s	Blackwell，Nvidia Grace Datacenter/Server CPU，NVSwitch for Blackwell

资料来源：维基百科，国信证券经济研究所整理

如果说 2016 年的 Pascal 架构在计算性能上有显著提升，同时支持英伟达的 NVLink 1.0，增强了 GPU 之间的通信能力是它的亮点，那么 2017 年发布的 Volta 架构的卖点则是**引入了 Tensor Core，这是英伟达开发的一种专门硬件加速器**。它通过混合精度计算技术，结合使用 FP16 和 FP32 数据格式，实现了在保持模型精度的同时大幅提升计算效率的目标，同时，Tensor Core 专门设计用于加速矩阵乘法运算，这是深度学习中最常见的操作之一。Tensor Core 的出现使得深度学习模型的训练和推理速度得到了显著提升，特别是在处理大规模数据集和复杂模型时表现出色。此外，Volta 架构支持更高的内存带宽，特别是通过 NVLink 2.0 和 HBM2，显著提升了数据传输速度。

2018 年英伟达发布了 Turing 架构，Turing 架构引入了 **RT Core，实现了实时光线追踪**，显著提高了图形渲染的质量和真实性，这是它最大的特色。同时，Turing 架构对 Tensor Core 进行了增强，支持 AI 推理和训练，加速了深度学习任务。这些进步使得 Turing 架构在图形渲染和 AI 计算领域都取得了显著的成果，为后续的架构创新奠定了基础。

图25：实时光影效果

资料来源：英伟达，国信证券经济研究所整理

图26：未加实时光影效果

资料来源：英伟达，国信证券经济研究所整理

2016 年开始，按照我们的划分，时代从移动互联网进入到人工智能时期的上半场，而上半场显著的特征是云计算的兴起。当时二级市场非常靓丽的企业是亚马逊，因为它的云计算以 50%左右的年化增长率在高速发展。因此，在这一时期，作为已经积累的丰富产品的供应商英伟达，其数据中心的业务开始崭露头角，且为公司的主营收入带来了可观的增长。

下表可见，英伟达数据中心产品从 2016 财报（2015 年）的 33.9 亿美元增长到 2020 财报（2019 年）的 298 亿美元，4 年复合增速高达 72%，成为公司产品线中当之无愧的最重要的增长点。

此外，在这一时期，公司的智能驾驶 GPU 也获得了较快的增长，复合增速达到了 21.6%，是增速第二高的产品线。

表11：英伟达 2015 年之后的各板块收入（百万美元）

自然年	2015	2016	2017	2018	2019	2020	2021	2022	2023	2015-2020	2015-2023
财报年	2016	2017	2018	2019	2020	2021	2022	2023	2024	CAGR	CAGR
游戏 GPU 和相关产品	281800	406000	551300	624600	551800	775900	1246200	906700	1044700	18.3%	17.8%
专业视觉设计 GPU	75000	83500	93400	113000	121200	105300	211100	154400	155300	12.7%	9.5%
智能驾驶 GPU 及相关产品	32000	48700	55800	64100	70000	53600	56600	90300	109100	21.6%	16.6%
数据中心产品	33900	83000	193200	293200	298300	669600	1061300	1500500	4752500	72.2%	85.5%
其他业务	78300	69800	77700	76700	50500	63100	116200	45500	30600	(10.4%)	(11.1%)

资料来源：wind，国信证券经济研究所

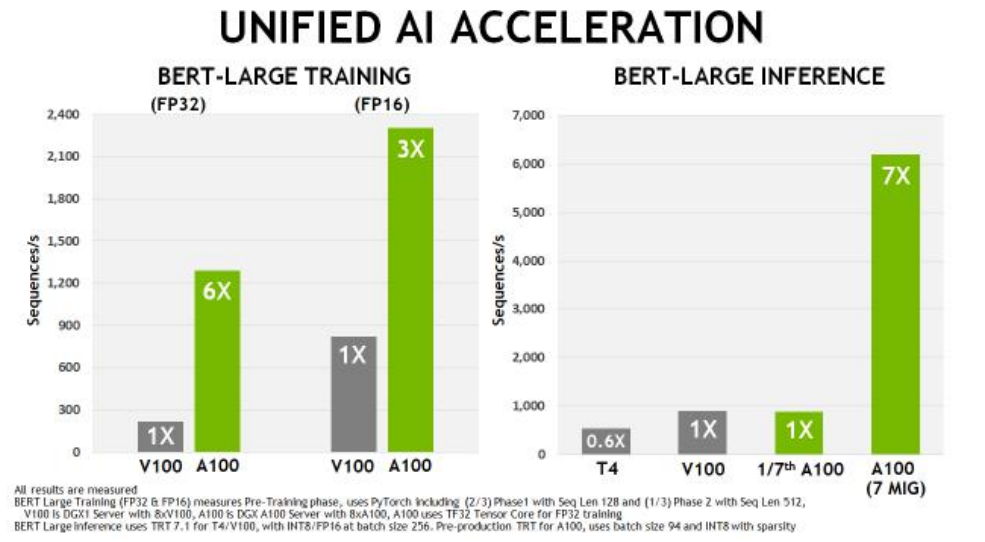
摩尔定律的延续者（2020 年-今）

2020 年，英伟达最重要的架构之一 Ampere 架构发布了。它引入了第二代 RT Core 和第三代 Tensor Core，大幅增强了光线追踪和人工智能计算能力，对于游戏和专业应用中的实时渲染、物理模拟和 AI 推理有着显著提升。基于 Ampere 架构的 A100 GPU 实现的第三代 NVIDIA 高速 NVLink 互连和新的 NVSwitch 显著增强了多 GPU 的可扩展性、性能和可靠性。第三代 NVLink 的数据速率为 50 Gbit/秒，是 V100 的 2 倍。

A100 还发布了一种新的架构 MIG（Multi-Instance GPU），这是一种硬件虚拟化技术，它允许将单个 GPU 划分为多个独立的 GPU 实例。每个实例都拥有自己的高带宽显存、缓存和计算核心，从而可以在单个 GPU 上并行运行多个工作负载，如推理、训练和 HPC 等，同时保持延迟和吞吐量的稳定性。

当时谷歌的 BERT 模型的知名度更高，甚至 OPENAI 也以 BERT 模型作为标杆企业来对比，因此 A100 对比了在 BERT 模型上的训练与推理速度。它在 FP32 精度下是 V100 训练速度的 6 倍，借助 MIG 技术在推理上 A100 速度可提升 7 倍。与 CPU 相比，在 BERT 等先进的对话式 AI 模型上，A100 可将推理吞吐量提升 249 倍。

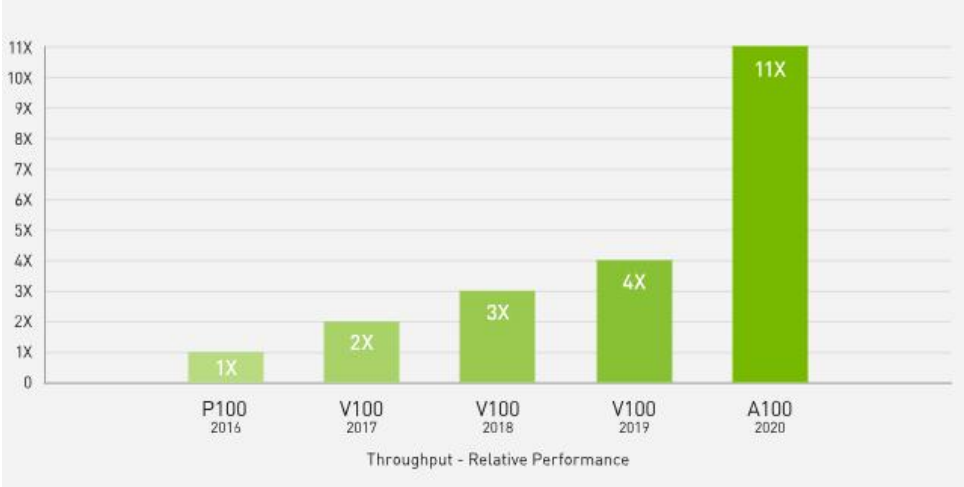
图27：BERT 模型训练与推理比较



资料来源：英伟达，国信证券经济研究所整理

此外，HPC（高性能计算）性能是指计算机系统在执行复杂、高负载计算任务时的表现。应用场景主要是大量计算资源的科学和工程领域，如天气预报、基因组学、流体力学仿真、材料科学、金融建模等。

图28：四年来 HPC 性能提升 11 倍

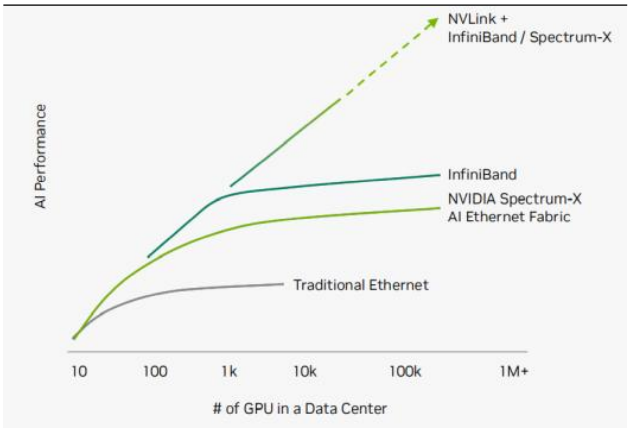


资料来源：英伟达，国信证券经济研究所整理

由于英伟达在 2019 年并购了 Mellanox，它的产品基于 InfiniBand 和以太网技术，Mellanox 为高性能计算、数据中心、云计算、计算机数据存储和金融服务等市场提供适配器、交换机、软件、电缆和硅片。因为以太网是有损网络，而 InfiniBand 是无损网络，InfiniBand 的速度天然比以太网速度更快。2022 年 400G 的 InfiniBand 产品发布，2024 年 800G 的 InfiniBand 产品发布，2024 年英伟达宣布 2025 年将发布 1600G 的 InfiniBand 产品。

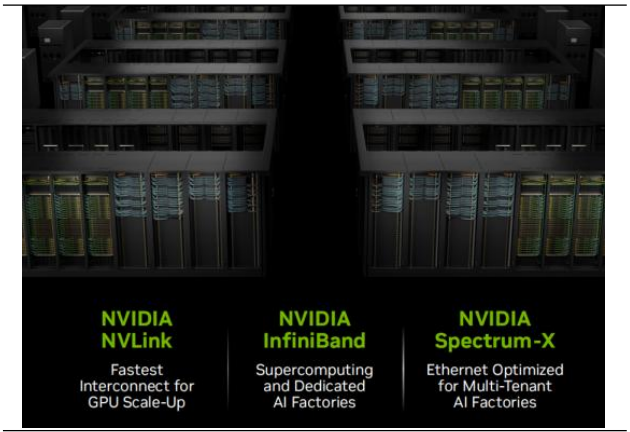
可见，此时的英伟达已经不再只关心 GPU 的速度，而是从平台层面上来审视“短木板”的每一个环节并逐一增强。包括 GPU 的 CUDA 核心、tensor 核心、光线追踪核心；网络侧的 NVLink，NVswitch，InfiniBand、Spectrum-X；以及软件堆栈与工具 CUDA 平台、cuDNN、cuBLAS、cuFTT 等，NVIDIA Deep Learning SDK，NVIDIA NGC；专用硬件包括 DGX 系列、Jetson 系列。

图29：生成式 AI 性能与网络技术的关系



资料来源：英伟达，国信证券经济研究所整理

图30：英伟达网络侧三大关键技术



资料来源：英伟达，国信证券经济研究所整理

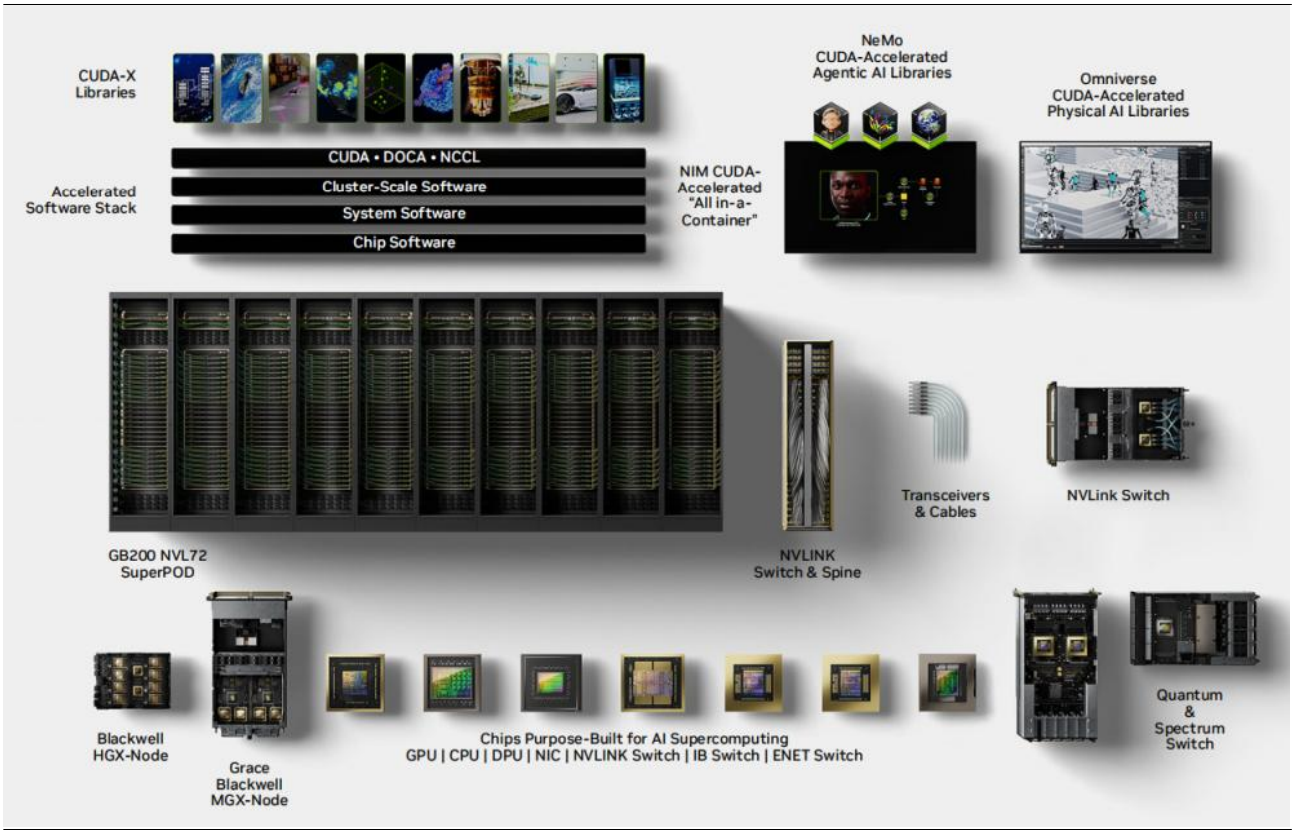
从此时开始，英伟达已经逐渐脱离了硬件公司的定位。英伟达应用深度学习研究

副总裁 Bryan Catanzaro 所说，“很多人不知道这一点，但 Nvidia 的软件工程师比硬件工程师还多。”

再系统点说，黄仁勋则认为：“公司的目标是建立一个架构，一个可以无处不在的平台；英伟达不是硬件公司，而是软件公司，更是个提供数据中心的全栈公司。”

下图可见，英伟达的加速平台，涉及到了芯片、网络、系统、软件和算法方面的创新。

图31：英伟达加速平台



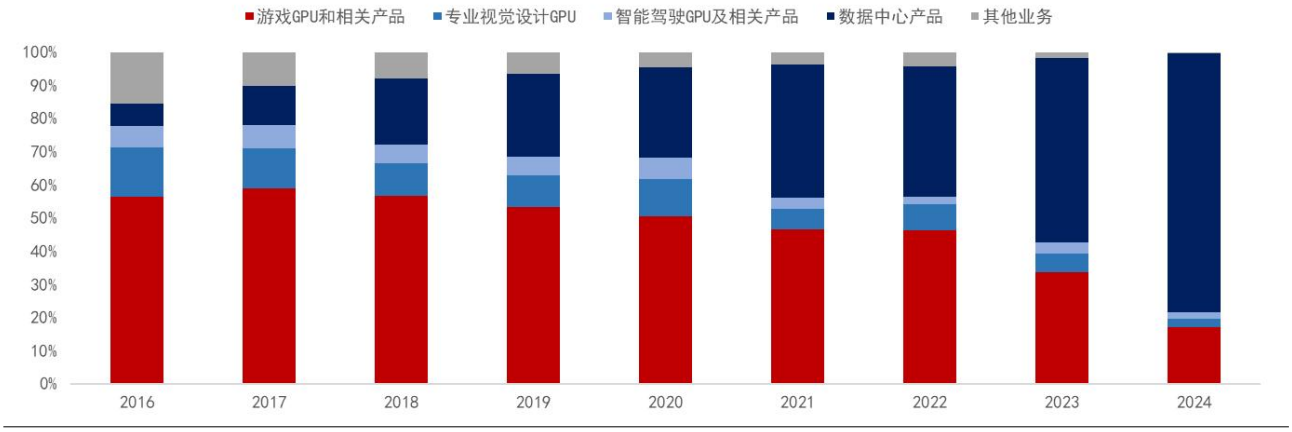
资料来源：英伟达，国信证券经济研究所整理

2020 年，正值全球的新冠疫情蔓延，大量的企业有办公需求，因此当年的电脑升级、数据中心扩容需求很大。加之 A100 的表现实在是惊艳，也获得了客户的大量订单，英伟达的各条业务线也是突飞猛进。其中，游戏 GPU 和相关产品同比增长了 40.6%，数据中心更是大幅增长了惊人的 124.5%！

在 2020 年-2021 年，市场多少分不清到底是新冠疫情带来的数据中心扩容需求，还是云计算或者 AI 驱动带来的扩容需求。回头来看，新冠疫情带来的需求在个人电脑产品侧只维持了两年（2020-2021 年），而由 AI 驱动的数据中心需求其实从 2020 年就开始发力了。

从那以后，英伟达的数据中心产品在收入中的占比一直提升、提升、再提升，到了 2024 财年，其占收比已经高达惊人的 78%！遥遥领先于其他产品线。

图32：2016-2024 财年英伟达的收入结构



资料来源：wind（英伟达财报发布在 1 月，财报年份-1 对应的是自然年，如 2016 年报对应的 2015 年），国信证券经济研究所整理

2022 年，公司发布了 Hopper 架构，Hopper 采用 4nm 工艺制造，拥有超过 800 亿个晶体管，核心产品是英伟达 H200 和 H100 Tensor Core GPU，并在生成式 AI 训练和推理方面实现了比上一代更高水平的加速。它有五项突破性创新：

1、**针对 Transformer 模型优化**：Hopper 架构通过 Transformer Engine 推进了 Tensor Core 技术，Hopper Tensor Core 能够应用混合 FP8 和 FP16 精度，从而显著加速 Transformer 的 AI 计算，将 TF32、FP64、FP16 和 INT8 精度的 FLOPS 提高了三倍；

2、**更快的网络**：第四代 NVLink 可以使用英伟达 DGX 和 HGX 服务器扩展多 GPU 输入和输出，每个 GPU 双向传输速度为 900 GB/s，是 PCIe Gen5 带宽的 7 倍多。第三代 NVSwitch 与上一代 A100 相比，在 8 个 H200 或 H100 GPU 服务器内可将吞吐量提高 2 倍。带有 NVLink 交换机系统的 DGX GH200 系统支持多达 256 个连接的 H200 集群；

3、**机密计算**：Hopper 架构推出了世界上第一个具有机密计算功能的加速计算平台。用户可以在本地、云端或边缘运行应用程序，并确保未经授权的实体在使用时无法查看或修改应用程序代码和数据；

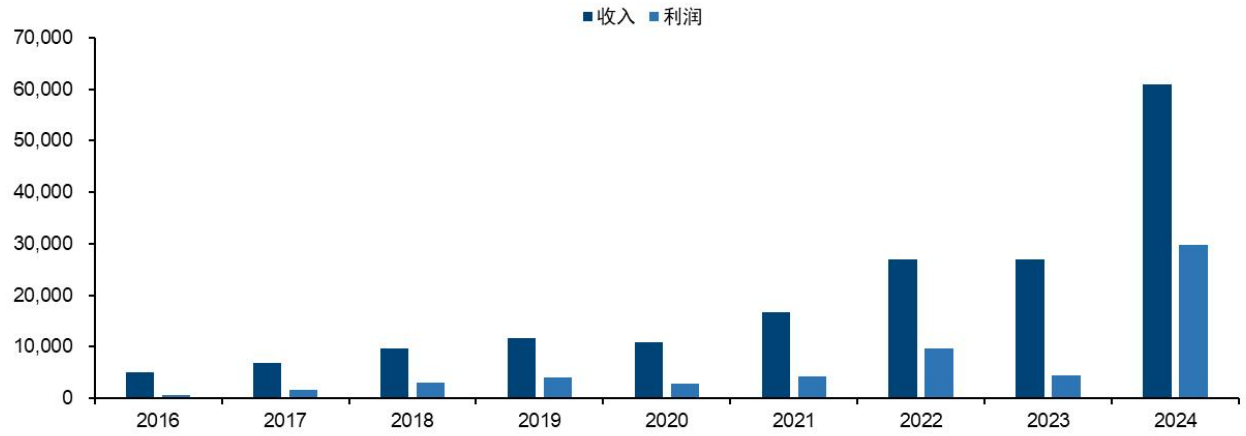
4、**第二代 MIG**:Hopper 架构通过在最多七个 GPU 实例的虚拟化环境中支持多租户、多用户配置。借助 Hopper 的并发 MIG 分析，管理员可以监控合适大小的 GPU 加速并优化用户的资源分配。对于工作量较小的研究人员，他们可以选择使用 MIG 来安全地隔离部分 GPU，而不是租用完整的 CSP 实例；

5、**动态规划**：与传统的双插槽 CPU 服务器相比，Hopper 的 DPX 指令可将动态规划算法的速度提高 40 倍，与 Ampere 架构相比，可将动态规划算法的速度提高 7 倍。这可显著加快疾病诊断、路由优化甚至图形分析的速度。

此外，2022 年末，正赶上 Chat GPT3.5 发布后产生了席卷全球的热度，AI 炙手可热，而 A100、H100 与 H200，成了无数巨头在资本开支中的首要选择。英伟达的收入更实现了爆发式的增长！

下图可见，英伟达从 2022 财年收入开始继续加速增长，其中 2023 财年的收入与 2022 财年相近，但结构差别很大，其中游戏 GPU 和相关产品线收入下滑 27.2%，但是数据中心产品线则增长了 41.4%，因此 2023 年总收入持平；2024 财年数据中心增长达到了惊人的 216.7%！2024 财年总收入也达到了 600 亿美元，利润更是逼近了 300 亿美元。

图33：2016-2024 财年英伟达的收入与利润（百万美元）



资料来源：Factset，国信证券经济研究所整理

2024 年 3 月，英伟达发布了 Blackwell 架构，Blackwell 架构 GPU 具有 2080 亿个晶体管，采用专门定制的 4nm 工艺制造。它的第二代 Transformer 引擎使得基于 Transformer 的大模型在训练上速度更快，在 FP4 精度下，其推理性能比 Hopper 提高了 30 倍，AI 性能比 Hopper 架构提高了 5 倍。网络侧的第五代 NVLink 也使得 GPU 之间的传送速度更快。

此外，市场预期 GB300 将在 2025 年 Q2 发布。它或将继续采用台积电 4 纳米工艺制程，同时针对计算芯片进行了优化设计，其算力性能相较 B200 可能再提升 50%。

如何看待英伟达的未来？

第一，英伟达是一家指数级思维的公司。

黄仁勋曾经说：市场的定义者（Market maker）从来不考虑市场份额。因此这就是为什么有人问黄仁勋：某某公司也在做 GPU，它们对你们有挑战吗？他总是回答：这是不一样的概念。别人做的是产品，而英伟达做的是平台。平台就像一个飞轮，英伟达审视并强化飞轮的每一个部分。

今年 62 岁的黄仁勋见过科网泡沫的疯狂，也经历过泡沫破裂后的低谷，他敏锐地捕捉到了开发人员使用 GPU 来做并行运算的需求，因此投入大量研发在 CUDA 的开发，即便遭遇市场的不理解和金融危机依然没有动摇他的想法。

在移动互联网机会来临时，公司也没有过格的 all in 移动端，而是有选择性地拓展，因为公司深知大功耗才是自己的优势。

提到摩尔定律，他的评价是：2000 年前后的 CPU 与软件是分离的，软件应用企业等待着 CPU 的突破，然后软件再跟上。而 GPU 则不同，算力的表现本身有硬件上的努力，更有软件、算法、流程、API 等各方面的进步。甚至人们将这些主张称之为“黄氏定律”——即，GPU 将推动 AI 性能实现逐年翻倍。

在这样的背景下，英伟达对并行运算、AI 运算发生在行业中的各种环境变化了如指掌并成竹在胸。他们思考的永远是：如果我们继续保持下一代架构能够在 AI 性能上提升 4 倍以上（假定 2 年一个新架构），那么目前制约这个飞轮最大的短板在哪里？如何解决这个短板？自研还是并购？如何将这些能力有机地整合起来？目前公司有 32000 人，黄仁勋期望未来英伟达员工人数能够突破 50000 人，而且他对 AI 是乐观的，他认为未来公司的 50000 人背后，可能是 1 亿个各种人工

智能助手在支撑，这样会给公司劳动生产率更大的提升空间。

黄仁勋在 2021 年和 2024 年被列入《时代》杂志年度 100 强榜单，这是《时代》杂志每年评选的全球 100 位最具影响力人物之一；2023 年 12 月被《经济学人》评为 2023 年最佳首席执行官；2024 年 2 月他因“高性能图形处理单元推动了人工智能革命”而当选美国国家工程院院士。

如前文所说，英伟达的核心竞争力早已不再是 GPU 本身，而是一种系统性的、平台性的核心竞争力。从硬件设计，到 CUDA 软件、库，到网络整合...简而言之，这是一种端到端的堆栈综合实力。因此，当将这些能力聚集到一块儿的时候，我们很难发现英伟达短期的对手在哪里，相信它在未来相当时间依然将领导算力革命。

第二，英伟达的天花板在哪里？

现在的 AI 竞争，如同一群探险家在沙漠中走了很久，突然间在遥远的天边看到了绿洲，请问，探险家什么反应？假如一个人对其他人说，你们等着，我去探探路，其他人多半会说，为什么不是我去探路，你们等着？

当通用人工智能（AGI）的梦想被点燃那一刻，就好比那充满生机的绿洲景象映入了探险家的眼里。探险家就像诸多科技巨头，有谁能抵制住沙漠中的绿洲（AGI）的巨大诱惑呢？

我们在报告《科技周期探索之七，2016-2030 年：通用人工智能时代的到来》中提及，2027-2029 年，是 OPENAI，马斯克，黄仁勋预测的 AGI 时代到来的大约时间。其路径是：如果 1 万亿参数（ChatGPT 4）的大模型代表了“聪明的高中生”，而一年半之后，大约是 2025 年底-2026 年初的 10 万亿级参数的大模型代表了“博士生”，那么再一年半之后的 100 万亿参数大模型，可能将数倍聪明于博士，同时模型参数也来到了人类神经突触量级（100 万亿-1000 万亿）。那么我们基本可以将那个时间认同为通用性人工智能时代的开始。

想想会发生什么？可能在“博士”水平下，人工智能就已经可以在千行百业辅助人类从事各种工作，而 AGI 时代，它的能力会到一种怎样的水平，涌现的能力会多到什么程度呢？这就是科技界目前的处境——隐约看到绿洲，但直到走近它之前，我们无法预测那里有什么！但这种憧憬在心里却如此躁动，如此热切！在这个情境里，不同的探险家的行为就像军备竞赛，他们都希望尽快到达绿洲。

我们来初步做个计算，假如端到端训练 AI 大模型的理论训练时间用公式来表示：

E_t = 8 * T * P / (n * X)

其中 E_t 为训练时长（秒），T 为训练数据的 Token 数，P 为模型参数量，n 为 GPU 的数量，X 则为每个 GPU 的算力。

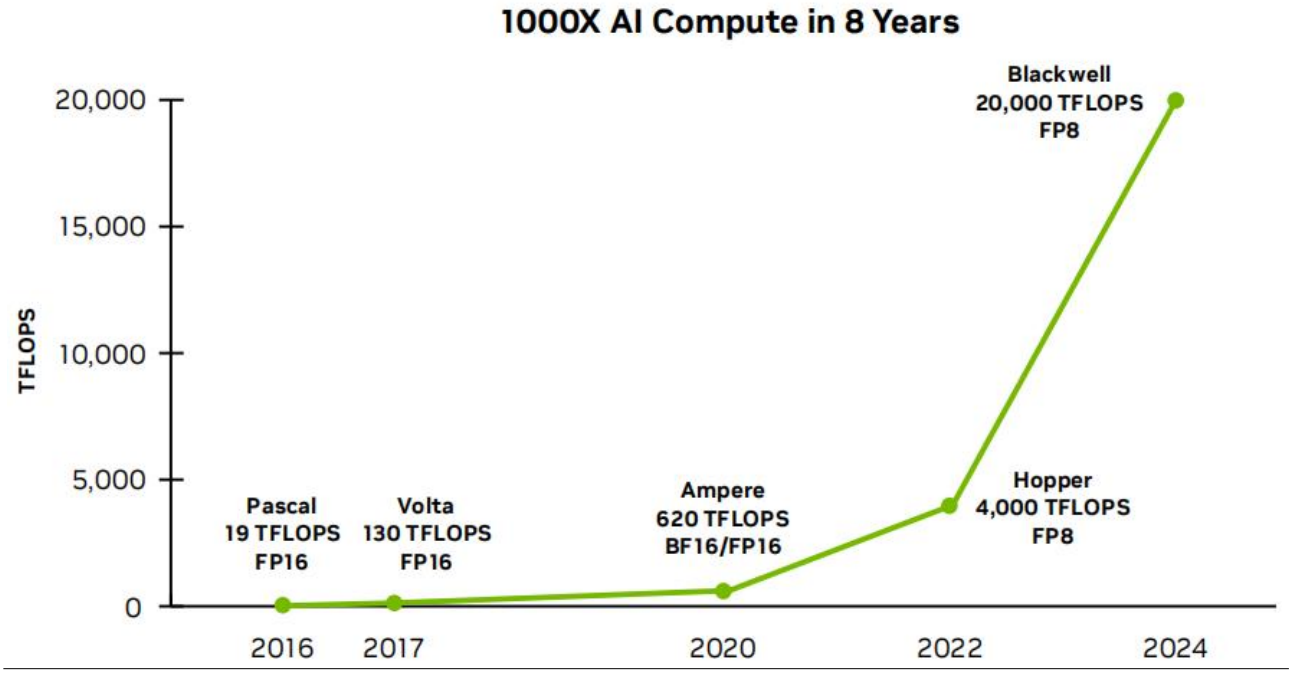
例如，Chat GPT 3.0，模型参数为 1750 亿，训练 token 大约为 3000 亿，按照 A100 GPU 理论最大 AI 计算性能 620 TFLOPS/s 的算力，训练该模型只要 100 块 A100，80 天的时间。

到了 Chat GPT 4.0，模型参数为 1.8 万亿，训练 token 大约到了 13 万亿，按照 H100 GPU 理论最大 AI 计算性能 4000 TFLOPS/s 的算力，训练该模型要 8000 块 H100，68 天的时间。当然，这是理论上的时间，实际上操作中可能有偏差，因此黄仁勋提到“训练一个 1.8 万亿参数量的 GPT 模型，需要 8000 张 Hopper GPU，消耗 15 兆瓦的电力，连续跑上 90 天”，这与我们理论上的测算相似。

倘若下一代大模型，参数是 10 万亿级的，其每个参数训练的 token 是 5-10 个（一般说来要训练到 20 个 token/每参数，模型达到较优的状态，但由于 Chat GPT 4.0 这个数字为 7.2，因此我们假定 5-10），则 8000 块 H100 跑下来时间就长得多了，要 1400-2800 天！若换成理论算力在 20Peta FLOPS（20000T FLOPS）的 Blackwell GPU，计算时间也要 280-560 天！如果我们认为 90 天是一个可以接受的训练时间的话，也就是说，30000-50000 块 GB200 才是能驾驭 10 万亿参数大模型的基础配置。

现在你可能意识到了，随着大模型参数翻上 10 倍，则对应的计算量大约翻 100 倍（在每个参数训练 token 数一样的情形下）。

图34：过去 8 年中，英伟达的 AI 算力翻了 1000 倍



资料来源：英伟达，国信证券经济研究所整理

因此，假定再下一代大模型，即我们之前讨论的相当于通用人工智能的 100 万亿参数问世时，倘若这个时间窗口依然在 3-4 年之后（平均 1.5-2 年迭代一个新量级大模型），即 2027-2028 年前后，我们需要的训练的运算量将是下一个版本的 100 倍，或者 Chat GPT 4.0 的 10000 倍。

按照英伟达 2024 年发布的 Blackwell 的 AI 算力较 Hopper 提升了 5 倍，假定 2026 年英伟达的下一个框架比 Blackwell 提升 4-5 倍，2028 年再下一代框架再提升 4-5 倍（大约维持目前的 AI 算力每年翻倍的能力），那么相较于 Chat GPT 4.0 所需的 GPU 的数量，也需要提升 10000/25=400 倍！也就是说，到了 2028 年，我们用当时最先进的英伟达 GPU，90 天完成一个百万亿参数大模型的训练，所需的 GPU 数量是 20 万块-40 万块。

或者说，假定英伟达 AI 算力提升的速度是每 2 年 5 倍的话，那么企业每 2 年所要购买的 GPU 数量将是此前的 20 倍！

2024 年 9 月，甲骨文的老板埃里克森提到：未来 4 到 5 年内，任何想参与这场大模型竞赛的企业，前沿模型门槛或高达 1000 亿美金，而且这场算力军备竞赛将永远进行下去。甲骨文最近宣布，将打造一个由 131072 个英伟达 GB200 NVL72

Blackwell GPU 组成的 Zettascale AI 超级集群 ，可提供 2.4 Zetta FLOPS 的 AI 性能，比马斯克的 xAI 算力集群更强大，后者目前拥有 100000 个英伟达 H100 GPU 显卡。

AMD 的 CEO 苏姿丰则表示，AI 芯片市场规模将以超过 60%的 CAGR 增长，并于 2028 年达到 5000 亿美元。

因此以此来看，在 AGI 没有实现之前，英伟达看不到天花板。

第三，英伟达的风险在哪里？

但问题是，如果 4 年之后，在 AI 芯片上的资本开支如果是 2024 年 4-5 倍的话，即便是微软、脸书这样的互联网巨头也会捉襟见肘。目前在它们的资本开支中，大约一半都投到了算力芯片上，如果在如此短的时间投资翻 4-5 倍的话（而它们的收入不能够也同步大幅增长的话），则无论是现金流，还是盈利能力，将无法承受如此之大的压力。

表12：部分公司财报季的资本开支（亿美元）

公司	2016	2017	2018	2019	2020	2021	2022	2023	2024
亚马逊(AMAZON)	67	101	113	127	350	554	583	481	
谷歌(ALPHABET)-A	102	132	251	235	223	246	315	323	
微软(MICROSOFT)	83	81	116	139	154	206	239	281	445
脸书(META PLATFORMS)	45	67	139	151	151	186	314	273	
苹果(APPLE)	135	128	133	105	73	111	107	110	
特斯拉(TESLA)	14	41	23	14	32	65	72	89	
甲骨文(ORACLE)	12	20	17	17	16	21	45	87	69

资料来源：wind，国信证券经济研究所整理

因此当我们再回到目前的情形：尽管 AI 芯片市场到 5000 亿美元似乎听起来不大，与目前全球智能手机市场规模 5000 亿美元大体相当，但由于能够参与到大模型建设的玩家太少（同每年十几亿部手机销量相比），即便如“探险家”般热情的企业也不得不考虑投入产出比。

相信在探索 AGI 的道路上，投资不会是一片坦途，**可能阶段性最大的敌人是经济周期的下行期**，一旦短期全球陷入滞胀，限于增长乏力与股东压力，企业的大规模投资必将会阶段性受阻，而英伟达的客户则面临：收入压力增加，被迫缩小资本开支，英伟达“因为客户的竞争所导致的大幅溢价”局面将会终止，从看着客户抢着买，到与客户商量着买，甚至是主动联系客户，则不同情境下的毛利率将会变化较大，届时收入增速放缓的英伟达可能会面临较大的市值波动风险。

但乐观来看，除了巨头们，随着 AI 芯片速度的提升，能够参与到千亿、万亿参数级别的大模型的门槛则将快速降低。目前万亿参数大模型对于大部分初创企业遥不可及，但到了 4 年之后，目前 8000 张 H100 算力卡的投入到时候变成了 400 张卡（1/20-1/25），这将是很多企业可以负担起的。而行业应用不一定需要 AGI 级别的大模型，而千亿、万亿级别大模型加上优化后的 Agent 智能体定会有广袤的舞台。

从这个角度说，一旦全球从滞胀周期中度过后，千行百业的人工智能应用依然将像雨后春笋般涌现。届时智能体、模型算力都会到一个更低的门槛水平，人工智能也不会像今天这样金字塔式的发展，由顶级大模型企业垄断着行业大部分能力，而应用企业起步的门槛过高，或许百花齐放的景象才是 AGI 时代真正的繁荣期！那个时候，AI 应用会更加扁平化，更加行业化，更加场景化，更加泛在化。

风险提示

地缘政治的不确定性，美联储降息幅度的不确定性，部分行业竞争格局的不确定性。

免责声明

分析师声明

作者保证报告所采用的数据均来自合规渠道；分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

国信证券投资评级

投资评级标准	类别	级别	说明
报告中投资建议所涉及的评级（如有）分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后 6 到 12 个月内的相对市场表现，也即报告发布日后的 6 到 12 个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。A 股市场以沪深 300 指数（000300.SH）作为基准；新三板市场以三板成指（899001.CSI）为基准；香港市场以恒生指数（HSI.HI）作为基准；美国市场以标普 500 指数（SPX.GI）或纳斯达克指数（IXIC.GI）为基准。	股票投资评级	优于大市	股价表现优于市场代表性指数 10%以上
		中性	股价表现介于市场代表性指数±10%之间
		弱于大市	股价表现弱于市场代表性指数 10%以上
		无评级	股价与市场代表性指数相比无明确观点
	行业投资评级	优于大市	行业指数表现优于市场代表性指数 10%以上
		中性	行业指数表现介于市场代表性指数±10%之间
		弱于大市	行业指数表现弱于市场代表性指数 10%以上

重要声明

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

国信证券经济研究所

深圳
深圳市福田区福华一路 125 号国信金融大厦 36 层
邮编：518046　总机：0755-82130833

上海
上海浦东民生路 1199 弄证大五道口广场 1 号楼 12 层
邮编：200135

北京
北京西城区金融大街兴盛街 6 号国信证券 9 层
邮编：100032