

「AI 巨头英伟达」的极端案例：受人工智能所驱动的美国股市前景

目前，英伟达(NVIDIA)及其图形处理器(GPU)客户成为推动股市回报、盈利增长、盈利修正、工业生产以及资本支出的重大因素。英伟达的财务业绩优异（公布的收益和盈利再次超出预期，并宣布了500亿美元的股份回购计划）。如下图所示，英伟达还是战后时期最快成为最大市值股票的公司。

但对于投资者来说，更重要的问题是，不光要考量销售 GPU 的经济效益，还要关注超大规模企业（谷歌、亚马逊、微软、Meta 等）以及其他人工智能基础设施用户/提供商从数以千亿美元计算的人工智能相关资本支出中赚取足够回报的能力。目前的支出水平可以媲美 1960 年代末的大型机时代和 1990 年代末大量铺设光纤的时期。为了在人工智能基础设施方面获得足够的回报，在未来 12 到 18 个月内，我们需要看到「推理」任务方面出现较大的利好发展（即使用人工智能(AI)运行企业客户的生产模型），而不仅仅是将 GPU 容量用于训练基础模型和聊天机器人。在本期《放眼市场》中，我们将进行深入探讨。

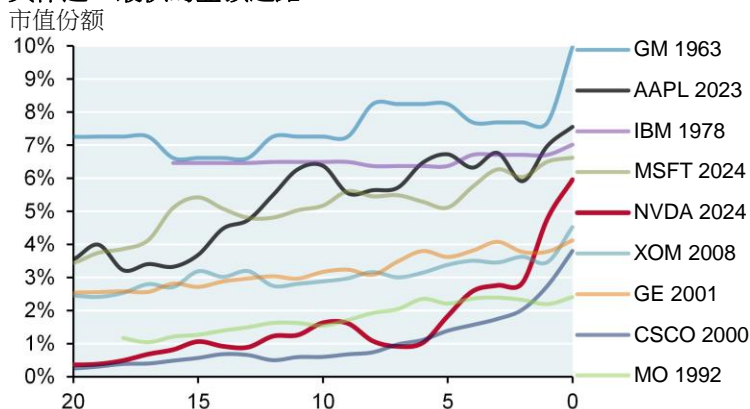
另外，我们还将简要评述第三巡回法院废除第 230 条有关某些社交媒体公司活动相关保护的決定。

岑博智先生(Michael Cembalest)
摩根资产管理

目录

- 「AI 巨头英伟达」：GPU 相关企业正在推动股票市场和资本支出2
- 英伟达的财务业绩：卓越非凡，与互联网时代的龙头企业截然不同4
- 从 A(Altman)到 Y(Combinator)：生成式人工智能的转型视角.....5
- 尽管如此……在支出、收入、生产力效益、能源等方面还存在很多疑问.....6
- 追踪迄今为止企业部门的人工智能采用情况：总体乐观，但存在一些好坏参半的信号.....8
- 股票市场在预期一些什么？9
- 「AI 巨头英伟达」预测：人工智能资本支出的「审判日」可能还剩 2-3 年时间10
- 附录一：英伟达目前和未来的竞争对手11
- 附录二：昔日股市龙头的命运.....12

英伟达：最快的登顶之路



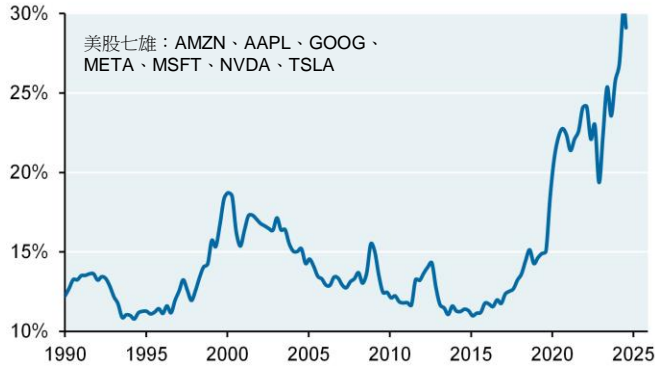
股票占据市值最大份额前季度
资料来源：彭博财经、Factset、GS、摩根资产管理，2024年8月

「AI 巨头英伟达」：GPU 相关企业正在推动股票市场和资本支出

英伟达及其 GPU 客户日益成为推动股市回报、盈利增长、盈利修正、工业生产以及资本支出的重大因素。您或许之前已看过下方第一张有关「美股七雄」市值份额的图表。其他图表同样值得留意，重点展示了盈利、工业生产和资本支出如何日渐受到人工智能相关支出的影响。最近，有同事指出，「美股七雄」的资本支出现在已超过整个能源行业。

美股七雄主导

占大盘股总市值的百分比



排名前五分之一的股票正在推动盈利修正和市场回报，百分比



美国高科技产业工业产值

指数 (100=2020年1月)



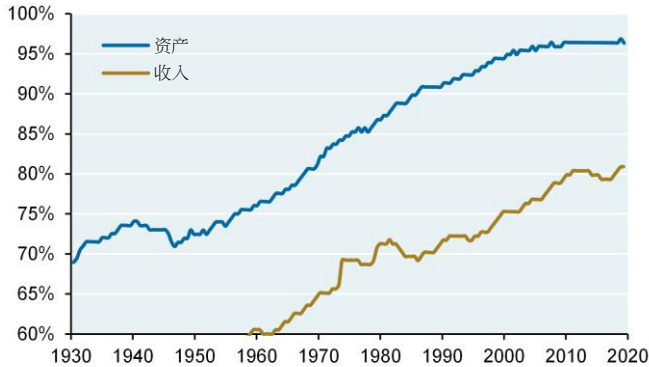
资本支出+研发，美股七雄与市场其他股票对比

指数 (100=2020年1月)



过去 100 年来，美国企业集中度一直在稳步上升；请参阅有关前 1% 公司所控制的资产和收入份额的图表。但我们现在遇到了新的极端情况，跑赢标普 500 指数的公司占比大幅下降，以及市值加权标普 500 指数跑赢等权重标普 500 指数（上一次发生这种情况是在 1990 年代末）就是例证。

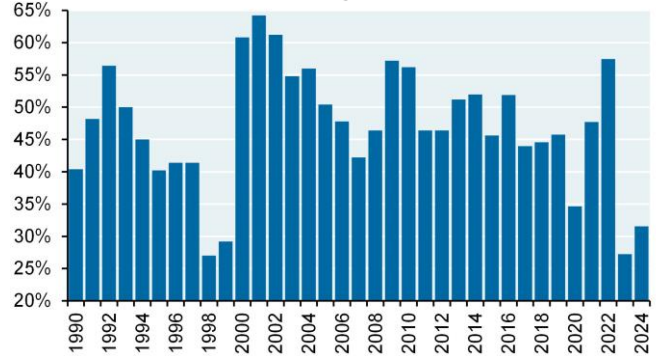
美国最大的1%公司的资产和收入份额，百分比



资料来源：「百年企业集中度」(100 years of corporate concentration)，Kwon 等人，2024年8月

跑赢标普500指数的成分股

表现跑赢大盘的公司百分比（总回报）



资料来源：彭博财经、摩根资产管理，2024年8月16日

市值加权相对等权重标普500指数的两年滚动表现



资料来源：彭博财经、摩根资产管理，2024年8月

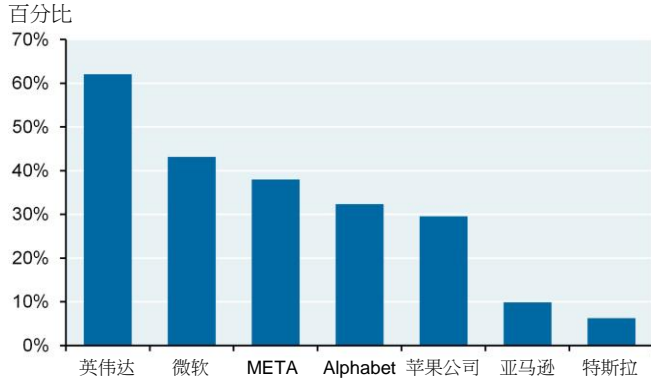
第三巡回法院废除第 230 条对社交媒体公司的部分保护

- 今年，在穆迪诉 NetChoice LLC 一案中，最高法院裁定，社交媒体或互联网公司使用算法以自动优先处理或以其他方式编排由第三方发布的内容，反映了该公司的编辑判断，构成受第一修正案保护的「表达活动」。换言之，您的 TikTok、推特或领英推送构成该公司自身的言论，尽管相关内容最初由第三方创建并上传。所以，法院的基本结论是，社交媒体公司享有第一修正案权利，以根据其所希望的任何政治、社会或经济标准，使用算法来提升或弱化帖文影响力。该裁定暗示，德克萨斯州最近通过的一项法律可能违宪，并将案件发回德克萨斯州下级法院审理
- 好，第三巡回法院表示：**如果经过编排的推送是第一方言论，那么《通信规范法案》第 230 条就不能使这些公司免于因其算法向用户推荐内容并造成损害而承担责任。**因此，在安德森诉 TikTok 案中，法院允许对 TikTok 提起诉讼。原告称，TikTok 的算法曾向一名儿童推送了一段「昏迷挑战」的视频，而该名儿童在进行该挑战时死亡
- 对我而言这一切都有道理；开始打官司吧

英伟达的财务业绩：卓越非凡，与互联网时代的龙头企业截然不同

- 英伟达的营业利润率很高，而且还在持续上升。去年，该公司的预期收入大幅增长（第三张图表，从金色点到红色点）。该公司发布了超预期的收入和盈利，并宣布了一项 500 亿美元的股份回购计划
- 英伟达的情况与思科（见下文）等互联网市场龙头截然不同，后者的市盈率虽也在飙升，但企业盈利却没有相匹配的增长
- 主流分析师报告预测，未来两年，英伟达将在人工智能芯片领域保持 90% 以上的市场份额。即使英伟达下一代「Blackwell」芯片的出货可能会延迟，但也没有什么公司能够立即取得优势，不过竞争正在到来。请参阅附录一——英伟达目前和未来的竞争对手及附录二——1960 年代以来股票在占据最大市值份额（譬如英伟达在 2024 年 6 月曾短暂称霸一天）后的走势

美股七雄营业利润率对比



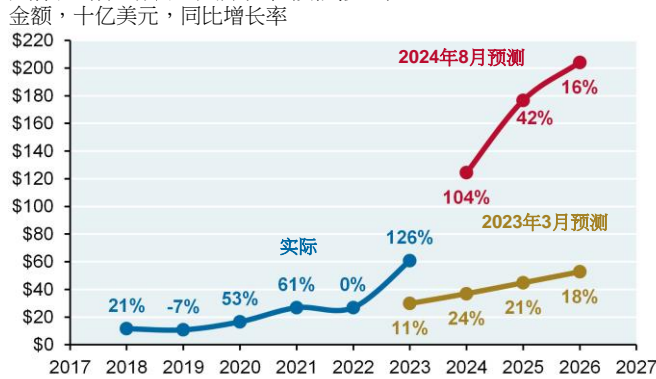
资料来源：彭博财经、摩根资产管理，2024年7月

英伟达净收入利润率



资料来源：彭博财经、摩根资产管理，2024年7月

英伟达销售增长：实际值和预估修正值



资料来源：彭博财经、摩根资产管理，2024年8月30日

市盈率，互联网时代与现在



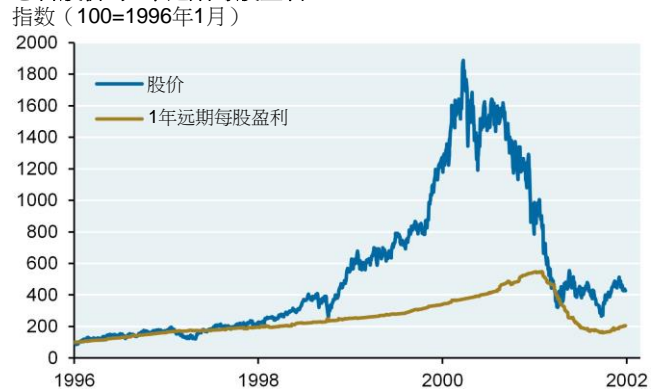
资料来源：Factset、摩根资产管理，2024年8月30日

英伟达股价与1年期预期每股收益



资料来源：Factset、彭博财经、摩根资产管理，2024年8月30日

思科股价与1年期远期每股盈利



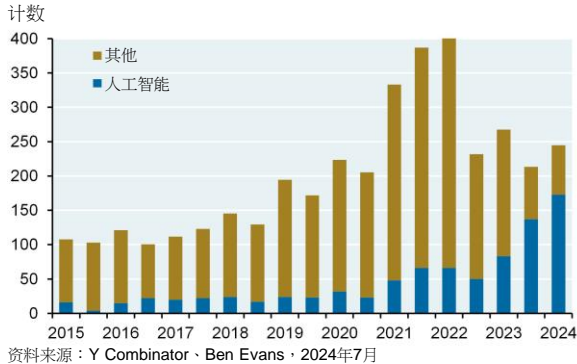
资料来源：Factset、彭博财经、摩根资产管理，2024年6月25日

从 A(Altman)到 Y(Combinator)：生成式人工智能的转型视角

「AI 巨头英伟达」的前景与其财务表现关系不大，而是与市场正在预期的人工智能转型是否/何时会真正发生有更大关系。人工智能的拥护者给出了强有力的理由：

- OpenAI 的山姆·奥特曼(Sam Altman)认为，人工智能是「一切技术革命中最大、最好且最重要的」，并坚信人工智能将逐渐变得更加无处不在¹
- 麦肯锡（该公司一向敢于做出轰动的预测，而这些预测往往不会有人去衡量是否准确）表示，生成式人工智能最终可能让全球经济每年增长约 8 万亿美元。²
- 大力促进风险投资的创投公司 Y Combinator 「全力押注」在人工智能领域之中，而其往绩表现卓越。贾里德·海曼(Jared Heyman)的 Rebel 基金估计了自 2005 年以来每笔 Y Combinator 交易的回报率。如果您对各笔交易都进行了投资（实际上并不可能），扣除摊销后的平均年回报率达到 176%³
- 人工智能正在推动风投生态系统，占 2024 年上半年新增独角兽企业的 40%以上，并占风投支持估值增长的 60%以上；迄今为止，美国独角兽企业的价值在 2024 年增长了 1,620 亿美元(Pitchbook)
- 人工智能的发展如此之快，以至于阅读理解、图像分类和高等数学的表现基准逐渐变得过时，取而代之的是各种新基准（见有关 GPQA 基准的文本框）。
- 生成式人工智能不仅可以提高呼叫中心、程序员和专业作家的效率；谷歌的一项研究预测，大约 80%的工作中至少有 10%的任务能够以快两倍的速度完成⁴
- 埃隆·马斯克(Elon Musk)新成立的 xAI 初创公司今年募集了 60 亿美元，让估值达到 240 亿美元。该公司计划在田纳西州建造全球最大的超级计算机，以支持人工智能训练和推理工作。马斯克通过分享由人工智能生成、内容模仿副总统哈里斯(Kamala Harris)的「深伪」视频，进一步加强其自由主义立场⁵
- 然后，阿森布伦纳(Aschenbrenner)发表了一篇长达 165 页改变思维的《态势感知》(Situational Awareness)文章，将人工智能的预测提升到了另一个层次（超级智能、美国电力供应的 20%、1 万亿美元以及到 2030 年实现全面的通用人工智能）⁶
- Radical Ventures 的 Rob Toews 将 AlphaFold 和其他蛋白质折叠算法描述为人工智能历史上最重要的成就（更多详情见 2024 年 4 月 2 日《放眼市场》第 3 页）。

Y Combinator按领域划分投资的初创公司



防谷歌问答基准(GPQA)

- 包含 400 多道多选题
- 即便是博士水平的学者，也只有 65%能够正确回答其领域内的问题。同样的一群学者，即使在不限使用互联网的情况下，在其研究领域之外的得分也只有 34%
- 2023 年，大部分人工智能模型在 GPQA 上得分 30%-40%。Claude 3（最新的 Anthropic 聊天机器人）得分约为 60%
- 然而……由于模型首先是按照测试模型的数据进行训练的，因此上述情况可能是模型过度记住和拟合答案所致

¹麻省理工人工智能，2024年5月

²《生成式人工智能的经济潜力：下一个生产力前沿》(The economic potential of generative AI: The next productivity frontier)，麦肯锡公司，2023年6月

³《论 YC 初创指数 176% 的年回报率，以及它为什么永远不会实现》(On the 176% annual return of a YC startup index...and why one will never exist)，贾里德·海曼，2023年7月。调查结果：5%-6%的 YC 初创公司成为价值 10 亿美元的独角兽企业，而这其中的 10%-12%成为了十角兽企业（十倍于独角兽企业）。不幸的是，不可能对每一笔 Y Combinator 交易都进行投资，因为在超额认购的初创轮中，各家公司都会亲自挑选投资者。此外，该 176% 的平均年回报集中在一小部分交易中。Y Combinator 的交易失败率很高——在前 17 批交易中，失败率约达 40%，而这还不包括陷入困境或被亏本出售的公司

⁴《生成式人工智能的经济影响》(The economic impact of generative AI)，Andrew McAfee, Visiting Fellow / Technology & Society，谷歌，2024年4月

⁵《灾难的根源：马斯克分享有关哈里斯的虚假人工智能视频后，专家发出警告》(Recipe for disaster: Experts issue warning after Musk shares fake AI video of Harris)，Salon.com，2024年7月

⁶在《态势感知：未来十年》(Situational Awareness: the Decade Ahead) (2024年6月)中，利奥波德·阿森布伦纳(Leopold Aschenbrenner)认为，他是世界上寥寥数百个真正理解人工智能转型范围和规模的人之一。他解释了数亿个人工智能系统如何使人工智能研究实现自动化，从而产生「超级智能」，而这将成为人类有史以来建造的最强大的武器……而如果中国首先实现真正的通用人工智能，西方世界将受威胁

尽管如此……在支出、收入、生产力效益、能源等方面还存在很多疑问

- OpenAI 今年是否真如 The Information 报道那样亏损 50 亿美元？⁷
- 根据 Anthropic 首席执行官引述的数据，到 2027 年，训练一个人工智能模型是否真的需要花费 1,000 亿美元？⁸
- 根据红杉资本大卫·卡恩(David Cahn)估计的数据，最大型的科技公司是否能够填平每年 5,000 亿美元的「收入缺口」来弥补人工智能数据中心的支出？

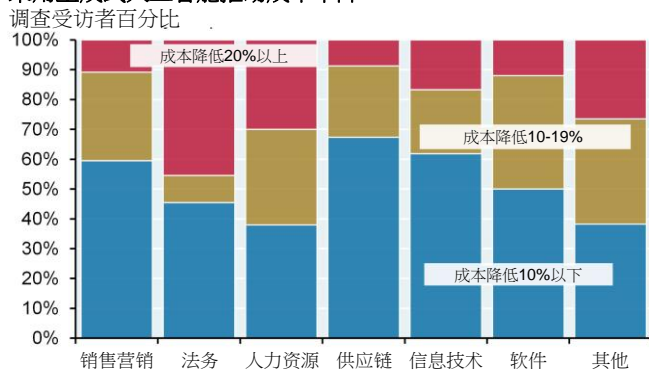
红杉资本：超大规模企业和企业应用软件公司是否会赚到足够的钱以支付人工智能基础设施？

卡恩将英伟达的收入乘以 2，得出数据中心的总成本，因为 GPU 大约占数据中心拥有成本的一半（其余是能源、建筑、发电机等）。然后，卡恩再将其乘以 2，以反映主要超大规模企业（即通过 Azure、亚马逊网络服务和谷歌云平台提供云服务的微软、亚马逊和谷歌）50%的预期毛利率。卡恩认为，谷歌、苹果、微软和 Meta 每年将创造 100 亿美元与人工智能相关的新增收入，而甲骨文、字节跳动、阿里巴巴、腾讯、X 和特斯拉将分别创造 50 亿美元。新增人工智能收入与人工智能数据中心支出之间的差额：每年 5,000 亿美元

《人工智能的 6,000 亿美元问题》(AI's \$600 billion question)，大卫·卡恩，红杉资本，2024 年 6 月

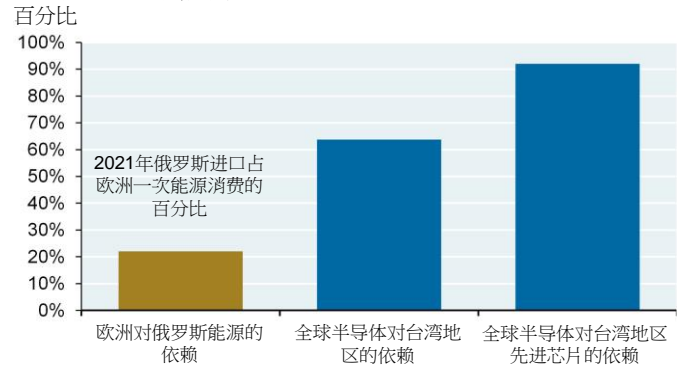
- 超大规模企业是否真在建造足够的人工智能计算基础设施以支持 12,000 个 ChatGPT（假设其目前的规模是每天 2 亿名活跃用户每天查询 1.25 个问题）？⁹ 巴克莱可能严重低估了广泛采用的生成式人工智能应用的查询强度和用户基础，特别是如果这些应用牵涉需要更多算力的查询线程……但是，您应该能抓住要点所在
- 巴克莱估计，在 2024 年，以最高使用率计算，已建成的 GPU 足以产生约 1,000 亿美元收入。GPU 用户在 2024 年的实际付款：约 100 亿美元。需要多长时间才能填补此差距？
- 为什么在 NewStreet Research 的调查中，大部分受访者表示，在生成式人工智能采用后，成本只会下降 10%或更少（见下文图表）？
- 为什么工程师使用人工智能编码工具的部分大型企业仍未看到端到端软件交付周期的改善？注：软件开发人员只花约 30%的时间在编码上
- 鉴于台积电目前是英伟达先进芯片的唯一供应商，市场何时开始多考虑地缘政治风险？¹⁰

采用生成式人工智能推动成本下降



资料来源：NewStreet Research，2024年8月

关键商品的区域依赖度



资料来源：英国石油公司统计评论，台湾地区，全球卫报，2024年

⁷ 《为什么 OpenAI 今年可能亏损 50 亿美元》(Why OpenAI Could Lose \$5 Billion This Year), The Information, Amir Efrati 和 Aaron Holmes, 2024 年 7 月

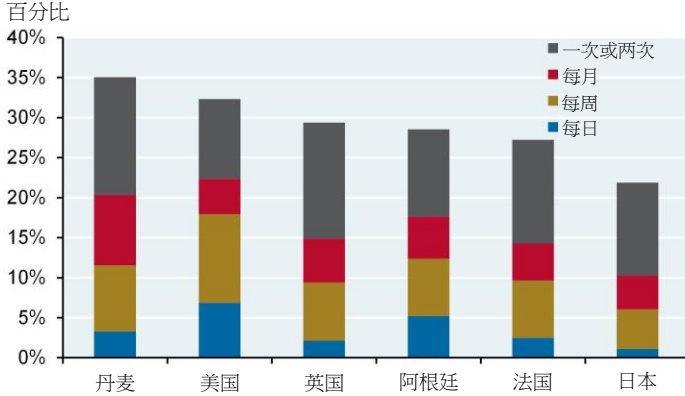
⁸ 《与客人 Dario Amodei 相处良好》(In Good Company with guest Dario Amodei), 挪威央行投资管理公司的播客, 2024 年 6 月

⁹ 《云人工智能资本支出：害怕错过还是梦幻成真？》(Cloud AI Capex: FOMO or Field-Of-Dreams?), Ross Sandler, 巴克莱股票研部, 2024 年 6 月

¹⁰ 截至 2023 年 8 月，英伟达通过台积电采购全部先进人工智能芯片。2024 年 7 月，英伟达宣布开始多元化供应来源，并开始也从三星采购先进存储芯片

- 如果麻省理工的 Daron Acemoglu 是对的，人工智能每年对全要素生产率的促进率只有 0.06%（鉴于过去 20 年全要素生产率的增长率平均每年达约 0.6%，这样的促进效果聊胜于无），那又如何？¹¹
- 为什么这么多 Chat GPT 用户像我一样不经常使用它？

ChatGPT的使用情况（按国家/地区划分）



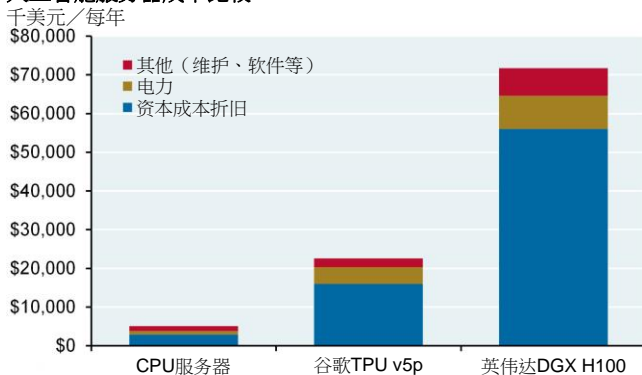
资料来源：路透社、Ben Evans，2024年5月

「因何而生」： GPT 和《放眼市场》

在 2023 年 9 月的《放眼市场》（《因何而生》）中，我解释了为什么我不会经常使用 GPT。我向 GPT-4 提出了 70 个出自《放眼市场》档案的问题，这些问题与市场、经济、政治和能源有关。GPT-4 的总得分是：C+。它答对了一半问题，但另一半答错了，尽管它能够进行网页检索、打开 PDF 等操作，然而它的错误基本上无法预测，迫使我必须检查每一个答案。如果情况属实，对我来说，任何生产力的增益都是有限的

- 那么能源消耗呢？就 Talen Energy 将其部分核电业务出售给亚马逊一事，早有争议。AEP 和 Exelon 向联邦能源管理委员会投诉，表示 1.4 亿美元的输电成本现在由 PJM 电力客户承担。请记住：云计算的出现几乎没有改变电力需求，因为数据中心取代了本地工作负载，而本地工作负载每单位计算的用电量比云计算高 2-3 倍。但生成式人工智能的崛起就没有这么好运了，它需要新的能量
- 来到 2024 年，人工智能/机器学习是否应该比它最近在检测乳腺癌方面的糟糕表现做得更好？人类医生发现了 18 例人工智能遗漏的乳腺癌，而人工智能只发现了 2 例人类医生遗漏的乳腺癌，并生成数百例假阳性¹²
- 英伟达推出的下一代芯片是否会降低人工智能服务器的成本？该公司声称，一个大型模型将能够使用 2,000 个每个成本 3.5 万至 4 万美元的 Blackwell GPU 训练，所需电力为 4 兆瓦，以取代需要电力达到 15 兆瓦的 8,000 个每个成本 2.5 万美元的 H100 GPU（即能够节省 50% 以上的成本和能源）。如果大量的推理任务可以由更便宜的 CPU 而不是 GPU 来执行，情况又将如何？

人工智能服务器成本比较



资料来源：NewStreet Research，2024年8月

美国核电复兴有多现实？



资料来源：动力堆系统数据库、摩根资产管理，2023年12月

¹¹Acemoglu (麻省理工) 首先假设生成式人工智能可将生产过程中约 20% 的任务转型，然后就以下估计做出调整：该数字只有 25% 能在 10 年内实现具有成本效益的自动化。他估计，与人工智能相关的劳动力成本节省约 25% 才能使估计的全要素生产率每年只提高 0.064%。《人工智能的简单宏观经济学》(The Simple Macroeconomics of AI)，Daron Acemoglu，麻省理工，2024 年 5 月

¹²《被人工智能检索和未检索的经筛查检测癌症的前瞻性分析》(A Prospective Analysis of Screen-Detected Cancers Recalled and Not Recalled by Artificial Intelligence)，Smith 等人，《Journal of Breast Imaging》，2024 年 5 月

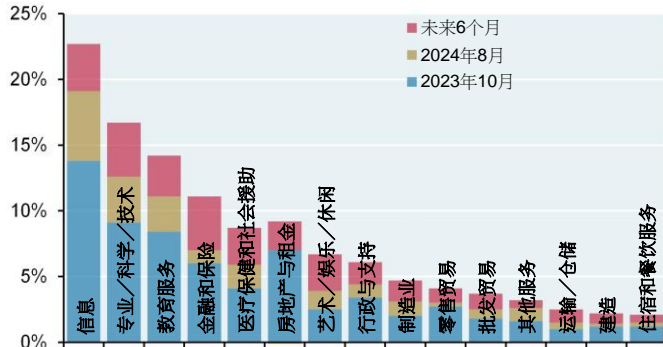
追踪迄今为止企业部门的人工智能采用情况：总体乐观，但存在一些好坏参半的信号

统计局追踪人工智能在其商业趋势和展望调查中的采用情况。鉴于从 2023 年 10 月到 2024 年 8 月人工智能的采用率稳步增长以及直至 2025 年的预测，调查结果是积极的（第一张图表）。首席执行官调查也表明，采用率正在上升：50%的受访首席执行官表示，他们正在招聘与生成式人工智能相关的职位（而这些职位去年还不存在），40%计划明年增加人工智能支出，70%表示生成式人工智能是他们的优先要务¹³。

但从贝恩的企业使用数据来看，人工智能的采用仍集中在开发/试点阶段；从 2023 年 10 月到 2024 年 2 月，实际生产使用案例略有下降。根据美联储数据¹⁴，人工智能工作技能的渗透率依然很低，这些数据更接近于 3D 打印早期阶段的情况，而不是像云计算和智能设备那样被迅速广泛采用（第三张图表）。在就业数据不断上升之际，科技行业就业人数下降是否表明人工智能的采用正在降低科技行业的劳动强度？目前尚不清楚；我们将不同类型的职位列表变化与对每种职位的人工智能颠覆性的估计进行了比较，发现只存在轻微的相关性。

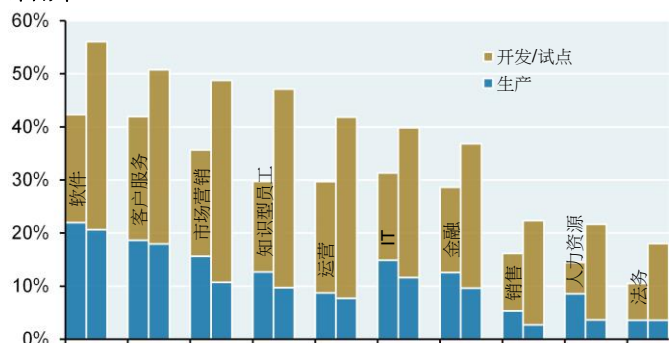
也许是我没什么耐心；20%的可寻址零售市场（不包括汽车和汽车零部件）耗时 20 年才转为以电子商务形式在网上销售，而人工智能的采用似乎会快一些。在 2007 年 iPhone 发布三年后，优步才出现，再过五年后，优步的月活跃用户才达到 2-3 千万。有一点是明确的：人工智能对咨询行业裨益颇巨。波士顿咨询公司目前有 20%的收入来自帮助大型公司研究如何利用生成式人工智能¹⁵。

人口普查：按行业和日期划分的人工智能采用率
使用人工智能的公司份额



资料来源：美国人口普查局、摩根资产管理，2024年8月

贝恩：2023年10月与2024年2月各阶段和行业的生成式人工智能采用率



资料来源：贝恩、Ben Evans，2024年7月

美联储：技术渗透到职位空缺
就业岗位显著的城市地区百分比



资料来源：圣路易斯联储、Kalyani & Hogan，2024年4月

美国私营部门就业



资料来源：美国劳工统计局、Haver、摩根资产管理，2024年7月

¹³IBM 人工智能首席执行官调查，2024 年 5 月；毕马威人工智能首席执行官调查，2024 年 4 月

¹⁴《人工智能与生产率增长》(AI and Productivity Growth)，圣路易斯联储储备银行，Kalyani 和 Hogan，2024 年 4 月

¹⁵《波士顿咨询公司表示，今年人工智能咨询将提供 20%的收入》(BCG says AI consulting will supply 20% of revenues this year)，《金融时报》，2024 年 4 月

股票市场在预期一些什么？

虽然英伟达股票的飙升伴随着企业盈利持续上升，但 38 倍的市盈率与股价处于历史上限区间的科技板块的其他股票一致。投资者态度乐观：科技股的空头头寸已经消失，市场适用于人工智能受益者的市盈率并不比适用于人工智能供应商的市盈率低多少。换言之，股票市场似乎已预期了大部分人工智能革命并在定价上有所反映。此外：指示性生成式人工智能 ETF¹⁶的总市盈率今年已升至 28 倍，不过仍低于 2022 年由流动性引发的 38 倍峰值。

至少现在的市盈率并不像以前科技热潮时期那样夸张：目前科技龙头的市盈率为整个市场的 1.8 倍，而互联网时代、大型机时代（1969 年）以及个人电脑时代（1983 年）的相对市盈率峰值分别达到 6 倍、2.8 倍和 2.4 倍¹⁷。

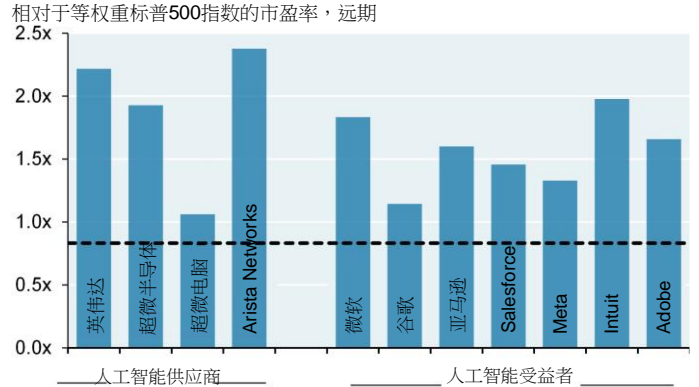
对投资者来说，好消息是：科技行业的资本支出和研发相对于现金流并没有上升，至少与 2000 年之前的时代相比时如此（第四张图表）。根据 Empirical Research，「美股七雄」的资本支出与收入的比率实际上低于整体市场，且预计这种情况会持续到 2026 年。

科技股和半导体股估值处于极端水平



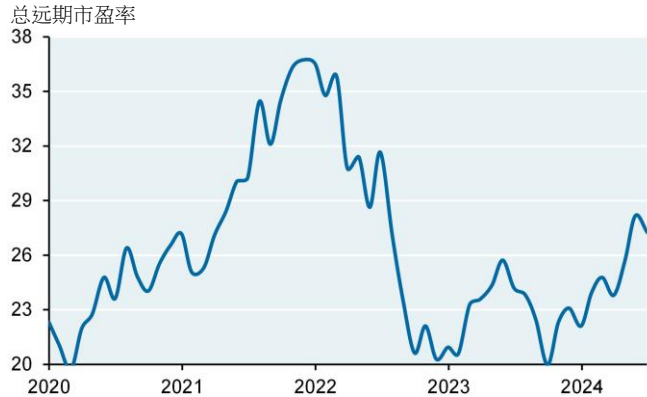
资料来源：彭博财经、摩根资产管理，2024年8月15日

人工智能供应商与受益者



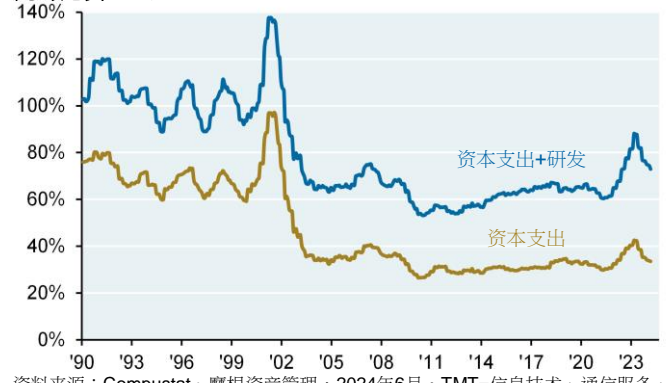
资料来源：彭博财经、摩根资产管理，2024年8月16日

指示性生成式人工智能ETF的估值



资料来源：彭博财经、摩根资产管理，2024年7月

标普500指数科技、媒体和电信(TMT)投资支出占经营活动现金流的比例，百分比



资料来源：Compustat、摩根资产管理，2024年6月。TMT=信息技术、通信服务、AMZN

¹⁶我们选择了一只以大权重配置英伟达（约为 7%）的人工智能 ETF。部分其他人工智能 ETF 今年表现欠佳，因为其英伟达权重低于标普 500 指数中的英伟达权重

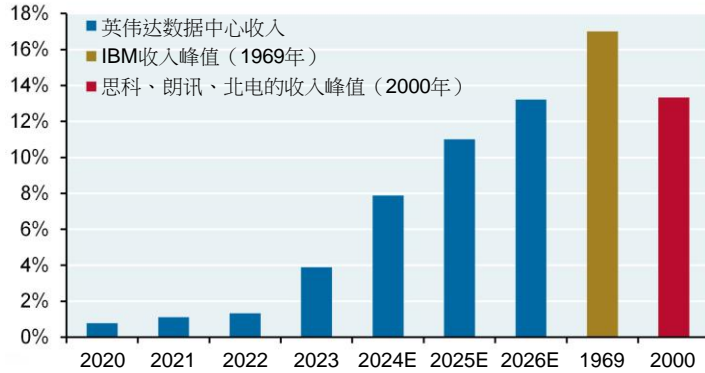
¹⁷《科技龙头：资本支出正在扼杀 CHAT》(The tech leaders: is capex killing the CHAT), Empirical Research, 2024 年 8 月 26 日

「AI 巨头英伟达」预测：人工智能资本支出的「审判日」可能还剩 2-3 年时间

- 每个计算周期首先需要基础设施，然后是平台，再然后才是应用。人们担心还没出现类似于 1990 年代的企业资源规划软件，或 2000 年代的搜索和电子商务应用等影响力重大的生成式人工智能应用，可能还为时过早
- 然而时间不会等人：英伟达数据中心收入占整个市场资本支出的份额预计将达到在 1969 年大型机时代和互联网繁荣时期所见的高峰（见图表）。因此，投资者需要承受很高的风险
- 下表显示，超大型企业和其他数据中心用户现在用在模型训练上的支出多于用在推理上的支出之情况¹⁸。简言之：这些公司用于训练复杂基础模型的支出是为最终客户日常运行完整生产模型所用支出的两倍以上。这种情况可能还会持续一年左右：盈利能力高的超大型企业将会保持人工智能资本支出机器的自主运转。此外：部分人工智能需求将来自军方，而出于国家安全考虑，需求的价格弹性通常较小
- 在未来两年内，第 6 页所示的企业人工智能采用趋势需要提高（即增加更多推理活动），以让所部署的全部资本避免产生「元宇宙」结果。如前所述，企业部门每年需要有数千亿美元的人工智能相关需求才能覆盖人工智能基础设施的成本

人工智能资本支出背景

占市场资本支出的份额



资料来源：实证研究，2024年8月

2024年人工智能数据中心支出

2024年数据中心运营成本

(单位： 十亿美元)	2024年人工智能数据中心支出			2024年数据中心运营成本		
	GPU和其他 芯片	其他人工 智能支出	人工智能资本 支出总额	培训 与研发	推理	总运营成本
微软	\$20	\$20	\$40	\$3	\$3	\$6
Meta	\$11	\$12	\$23	\$2	\$2	\$4
谷歌	\$14	\$15	\$29	\$3	\$1	\$4
亚马逊	\$8	\$8	\$16	\$2	\$1	\$3
二级云	\$26	\$26	\$52	\$8	\$3	\$11
企业及政府	\$26	\$26	\$52	\$8	\$2	\$10
总计	\$105	\$105	\$210	\$27	\$12	\$39

资料来源：New Street Research，2024年8月。运营成本包括现金运营费用、软件、折旧和电力

¹⁸行业经验法则：推理的成本约为训练成本的平方根

附录一：英伟达目前和未来的竞争对手

虽然英伟达目前在先进人工智能芯片上占有 90% 以上的市场份额，但有若干竞争对手目前提供或计划提供与英伟达 GPU 竞争的产品。

半导体公司

- AMD 已开发出与英伟达竞争的芯片，并开发出其自身的 ROCm 软件与英伟达的 CUDA 竞争
- 英特尔正在设计 Gaudi 3，其速度较英伟达 H100 芯片快 1.5 倍，而价格则便宜 30%-60%。英特尔还在与谷歌、高通等公司合作，以开发一款与英伟达 CUDA 软件竞争的产品。英特尔即将推出的 Granite Rapids 芯片将在每个核心上配备推理处理单元（适用于较小的模型），而 AMD 可能会跟随
- 软银旗下的 Arm 专注于设计和销售芯片 IP，但据报道，该公司正在与台积电商讨，以确保能够在 2025 年某个时候开始生产人工智能芯片

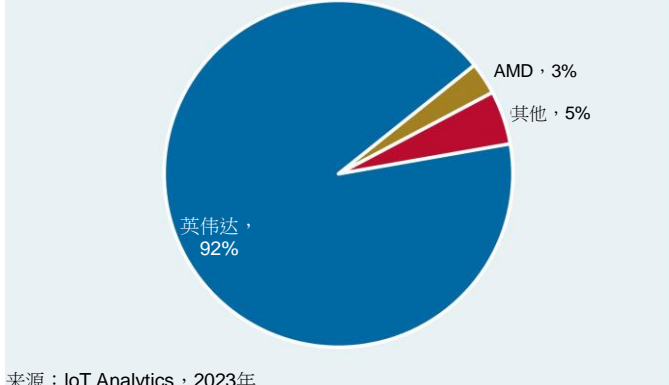
云服务提供商

- 亚马逊已经向现有的 AWS 客户提供 Trainium 和 Inferentia 芯片以及英伟达芯片。AWS 估计，利用自身芯片运行的模型成本大约是英伟达芯片的一半
- 谷歌已经向 GCP 客户提供 TPU v5p 芯片，年运行成本约为英伟达同类产品（H100 芯片）的三分之一。在一众云服务提供商中，谷歌是与英伟达最为接近的竞争对手
- 微软为 Azure 客户提供比 H100 更便宜的 Maia 100

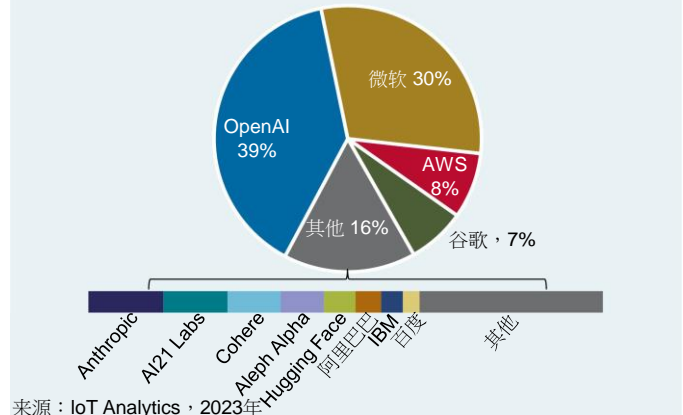
其他人工智能应用开发者和内部用户

- 苹果正在就内部推理运算推行一个名为 Project ACDC 的人工智能芯片项目
- Cerebras 与戴尔合作推销其 WSE-3 芯片，以与英伟达的 H100 芯片竞争。WSE-3 芯片具备 50 倍于 H100 的计算能力，能够训练规模相当于 GPT-4 10 倍的模型，但与向 AWS 购买计算能力类似，只能作为 Cerebras 计算平台的一部分购买
- 前谷歌 TPU 设计师创立的 Groq 正在开发一款经过优化的芯片，以实现针对计算强度较低的模型进行超快速推理。Groq 与 Global Foundries 而非台积电合作
- Meta 已为内部人工智能项目开发出其自营芯片——Meta 训练和推理加速器(MTIA)。MTIA 主要用于训练 Meta 的推荐和排名算法，但目标是让 MTIA 最终能够训练与 Meta 的 Llama 类似的生成式人工智能
- 特斯拉已在其 Dojo 超级计算平台上就自动驾驶汽车项目设计出其自营 D1 芯片
- 据报道，OpenAI 已开始就芯片开发与博通进行商讨，聘请了来自谷歌 TPU 团队的员工，以及正在筹集资金与英特尔/台积电/三星合作建设晶圆厂

根据2023年收入计算的数据中心GPU市场份额
百分比



基于2023年收入的人工智能模型和平台的市场份额，百分比

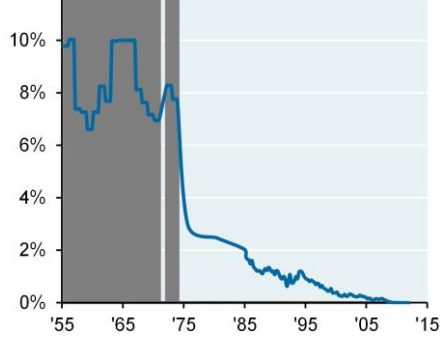


附录二：昔日股市龙头的命运

在 2024 年 6 月短暂的一日行情中，英伟达以 3.33 万亿美元的市值问鼎标普 500 指数，超过微软。过去几十年的市场龙头通常会到达某个拐点，其市值占市场的份额随后就会下降（通用汽车、IBM、奥驰亚、思科、通用电气和埃克森）。只有微软和苹果实现持续强劲的发展。注：灰色条表示各股票成为标普 500 指数市值最大的时期；市值份额按季度末基准计算

通用汽车

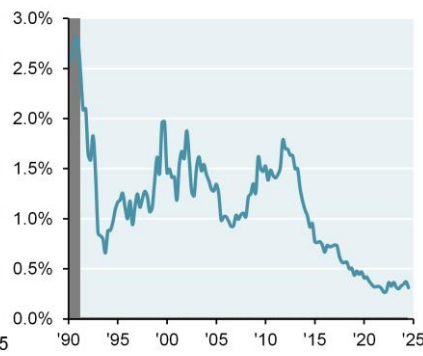
占总市值的份额



资料来源：Factset、Datastream、摩根资产管理，2024年6月

IBM

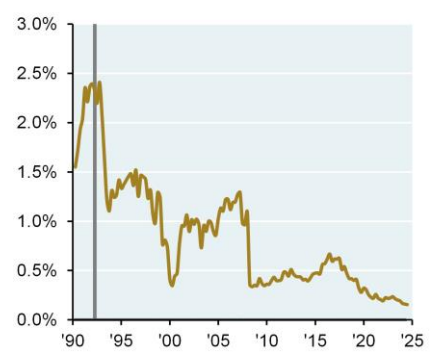
占总市值的份额



资料来源：Factset、摩根资产管理，2024年6月

Altria

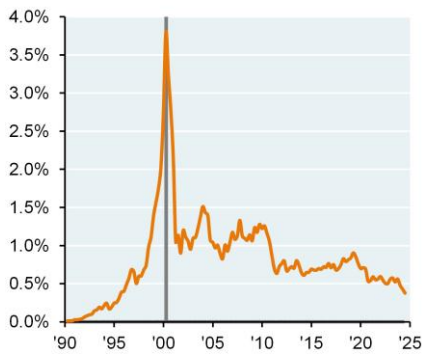
占总市值的份额



资料来源：Factset、摩根资产管理，2024年6月

思科

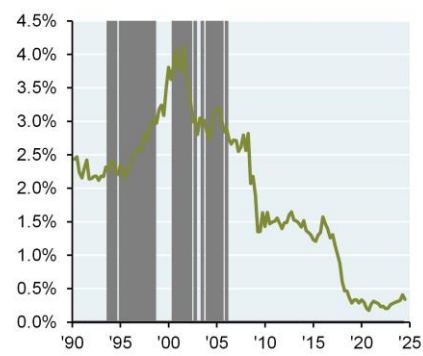
占总市值的份额



资料来源：Factset、摩根资产管理，2024年6月

通用电气

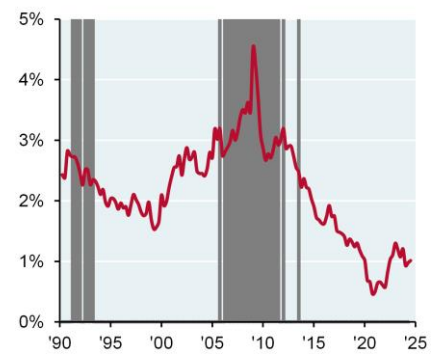
占总市值的份额



资料来源：Factset、摩根资产管理，2024年6月

埃克森美孚

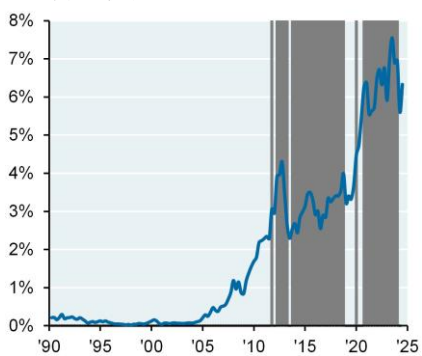
占总市值的份额



资料来源：Factset、摩根资产管理，2024年6月

苹果公司

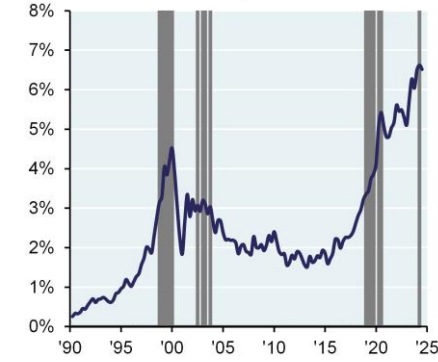
占总市值的份额



资料来源：Factset、摩根资产管理，2024年6月

微软

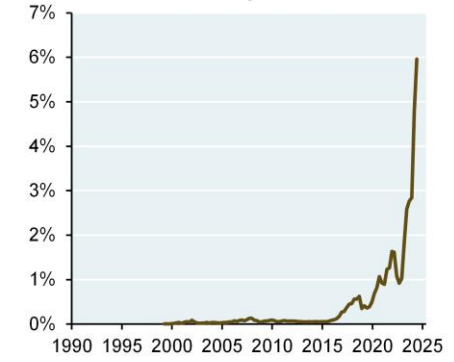
占总市值的份额



资料来源：Factset、摩根资产管理，2024年6月

英伟达

占总市值的份额



资料来源：Factset、摩根资产管理，2024年6月

重要信息

本文件仅供参考用途。本文件表达的观点、意见及预测，均为岑博智先生按目前市场状况作出的判断；如有更改，恕不另行通知，且可能与摩根大通的其他领域所表达的观点、意见及预测不同。**本文件不构成亦不应视为摩根大通研究报告看待。**文中提及的公司仅供说明用途而列示，不应视作摩根大通的建议或认可。

一般风险及考虑因素

本文件讨论的观点、策略或产品未必适合所有客户，可能面临投资风险。**投资者可能损失本金，过往表现并非未来表现的可靠指标。**资产配置/多元化不保证录得盈利或免招损失。本文件所提供的资料不拟作为作出投资决定的唯一依据。投资者务须审慎考虑本文件讨论的有关服务、产品、资产类别（例如股票、固定收益、另类投资或大宗商品等）或策略是否适合其个人需要，并须于作出投资决定前考虑与投资服务、产品或策略有关的目标、风险、费用及支出。请与您的摩根大通团队联络以索取这些资料及其他更详细信息，当中包括您的目标/情况的讨论。

非依赖性

本公司相信，本文件载列的资料均属可靠；然而，摩根大通不会就本文件的准确性、可靠性或完整性作出保证，或者就使用本文件的全部或部分内容引致的任何损失和损害（无论直接或间接）承担任何责任。我们不会就本文件的任何计算、图谱、表格、图表或评论作出陈述或保证，本文件的计算、图谱、表格、图表或评论仅供说明/参考用途。本文件表达的观点、意见、预测及投资策略，均为本公司按目前市场状况作出的判断；如有更改，恕不另行通知。摩根大通概无责任于有关资料更改时更新本文件的资料。本文件表达的观点、意见、预测及投资策略可能与摩根大通的其他领域、就其他目的或其他内容所表达的观点不同。**本文件不应视为研究报告看待。**任何预测的表现和风险仅以引述的模拟例子为基础，且实际表现及风险将取决于具体情况。前瞻性的陈述不应视为对未来事件的保证或预测。

本文件的所有内容不构成任何对您或对第三方的谨慎责任或与您或与第三方的咨询关系。本文件的内容不构成摩根大通及/或其代表或雇员的要约、邀约、建议或咨询（不论财务、会计、法律、税务或其他方面），不论内容是否按照您的要求提供。摩根大通及其关联公司与雇员不提供税务、法律或会计意见。您应在作出任何财务交易前咨询您的独立税务、法律或会计顾问。

就摩根资产管理客户而言：

「摩根资产管理」是摩根大通及其全球关联公司从事资产管理业务的品牌名称。

在适用法例所容许的范围内，我们可进行电话录音及监察电子通讯记录，藉以遵从我们的法律及监管规例和内部政策。摩根资产管理将会根据我们的隐私政策收集、储存及处理个人资料（详情可浏览：<https://am.jpmorgan.com/global/privacy>）。

可访问性

仅适用于美国：如果您是残障人士并需取得额外支援以查阅本文件，请致电我们寻求协助（电话：1-800-343-1113）。

本通讯文件由下列实体发行：

在美国，由摩根大通投资管理有限责任公司(J.P. Morgan Investment Management Inc.)或摩根大通另类资产管理有限责任公司(J.P. Morgan Alternative Asset Management, Inc.)发行，两家公司均须受美国证券交易委员会监管；在拉美，由当地摩根大通实体（视情况而定）发行并仅供指定收件人使用；在加拿大，由摩根资产管理（加拿大）有限责任公司(JPMorgan Asset Management (Canada) Inc.)发行并仅供机构客户使用，该公司乃加拿大所有省份及地区的已注册投资组合经理及获豁免市场交易商（除了育空），同时也是卑诗省、安大略省、魁北克省以及纽芬兰和拉布拉多等地的已注册投资基金经理。在英国，由摩根资产管理（英国）有限公司(JPMorgan Asset Management (UK) Limited)发行，该公司须受英国金融行为监管局授权及监管；在其他欧洲司法管辖区，由摩根资产管理（欧洲）有限责任公司(JPMorgan Asset Management (Europe) S.à r.l.)发行；在亚太地区，由以下发行实体在其主要受监管的司法管辖区区内发行：摩根资产管理（亚太）有限公司(JPMorgan Asset Management (Asia Pacific) Limited)，或摩根基金（亚洲）有限公司(JPMorgan Funds (Asia) Limited)，或摩根实物资产管理(亚洲)有限公司(JPMorgan Asset Management Real Assets (Asia) Limited)发行，各自均受香港证券及期货事务监察委员会监管；摩根资产管理（新加坡）有限公司(JPMorgan Asset Management (Singapore) Limited)（公司注册编号：197601586K），本广告或公告未经新加坡金融管理局审阅；摩根证券投资信托股份有限公司(Jpmorgan Asset Management (Taiwan) Limited)；摩根资产管理（日本）有限公司(JPMorgan Asset Management (Japan) Limited)，该公司乃日本投资信托协会(Investment Trusts Association of Japan)、日本投资顾问协会、第二类金融工具商同业公会及日本证券业协会的会员，须受日本金融管理局监管（注册编号：330(Kanto Local Finance Bureau (Financial Instruments Firm)）；在澳大利亚，由摩根资产（澳大利亚）有限公司(JPMorgan Asset Management (Australia) Limited (ABN 55143832080) AFSL 牌照号码：376919)，仅供按照公司法第 2001 第 761A 条及第 761G 条（《公司法》）赋予的定义的「批发客户」发行。在亚太所有其他市场，则仅向指定收件人发行。

就摩根大通私人银行客户而言：

可访问性

摩根大通一直致力于为所有客户提供符合其金融服务需要的产品及服务。如有任何关于产品及服务方面的问题，请致电摩根大通私人银行客户服务中心与我们直接联系（电话：1-866-265-1727）。

法律实体、品牌及监管信息

在美国，银行存款账户及相关服务（例如支票、储蓄及银行贷款）乃由**摩根大通银行(JPMorgan Chase Bank, N.A.)**提供。摩根大通银行是美国联邦存款保险公司的成员。

在美国，投资产品（可能包括银行管理账户及托管）乃由**摩根大通银行(JPMorgan Chase Bank, N.A.)**及其关联公司（合称「**摩根大通银行**」）作为其一部分信托及委托服务而提供。其他投资产品及服务（例如证券经纪及咨询账户）乃由**摩根大通证券(J.P. Morgan Securities LLC)**（「**摩根大通证券**」）提供。摩根大通证券是**金融业监管局**和**证券投资者保护公司**的成员。保险产品是透过 Chase Insurance Agency, Inc（「CIA」）支付。CIA 乃一家持牌保险机构，以 Chase Insurance Agency Services, Inc.的名称在佛罗里达州经营业务。摩根大通银行、摩根大通证券及 CIA 均为受 JPMorgan Chase & Co.共同控制的关联公司。产品不一定于美国所有州份提供。

在德国，本文件由**摩根大通有限责任公司(J.P. Morgan SE)**发行，其注册办事处位于 Taunustor 1 (TaunusTurm), 60310 Frankfurt am Main, Germany am Main，已获德国联邦金融监管局（Bundesanstalt für Finanzdienstleistungsaufsicht，简称为「BaFin」）授权，并由 BaFin、德国中央银行（Deutsche Bundesbank）和欧洲中央银行共同监管。在**卢森堡**，本文件由**摩根大通有限责任公司卢森堡分行**发行，其注册办事处位于 European Bank and Business Centre, 6 route de Treves, L-2633, Senningerberg, Luxembourg，已获德国联邦金融监管局（BaFin）授权，并由 BaFin、德国中央银行和欧洲中央银行共同监管。摩根大通有限责任公司卢森堡分行同时须受卢森堡金融监管委员会（CSSF）监管，注册编号为 R.C.S Luxembourg B255938。在**英国**，本文件由**摩根大通有限责任公司伦敦分行**发行，其注册办事处位于 25 Bank Street, Canary Wharf, London E14 5JP，已获德国联邦金融监管局（BaFin）授权，并由 BaFin、德国中央银行和欧洲中央银行共同监管。摩根大通有限责任公司伦敦分行同时须受英国金融市场行为监管局以及英国审慎监管局监管。在**西班牙**，本文件由**摩根大通有限责任公司 Sucursal en España (马德里分行)**分派，其注册办事处位于 Paseo de la Castellana, 31, 28046 Madrid, Spain，已获德国联邦金融监管局（BaFin）授权，并由 BaFin、德国中央银行和欧洲中央银行共同监管。摩根大通有限责任公司马德里分行同时须受西班牙国家证券市场委员会（Comisión Nacional de Valores，简称「CNMV」）监管，并已于西班牙银行行政注册处以**摩根大通有限责任公司分行**的名义登记注册，注册编号为 1567。在**意大利**，本文件由**摩根大通有限责任公司米兰分行**分派，其注册办事处位于 Via Cordusio, n.3, Milan 20123, Italy，已获德国联邦金融监管局（BaFin）授权，并由 BaFin、德国中央银行和欧洲中

央银行共同监管。摩根大通有限责任公司米兰分行同时须受意大利央行及意大利全国公司和证券交易所监管委员会 (Commissione Nazionale per le Società e la Borsa, 简称为「CONSOB」) 监管, 并已于意大利银行行政注册处以摩根大通有限责任公司分行的名义登记注册, 注册编号为 8076, 其米兰商会注册编号为 REA MI 2536325。在**荷兰**, 本文件由**摩根大通有限责任公司阿姆斯特丹分行**分派, 其注册办事处位于 World Trade Centre, Tower B, Strawinskylaan 1135, 1077 XX, Amsterdam, The Netherlands。摩根大通有限责任公司阿姆斯特丹分行已获德国联邦金融监管局 (BaFin) 授权, 并由 BaFin、德国中央银行和欧洲中央银行共同监管。摩根大通有限责任公司阿姆斯特丹分行同时须受荷兰银行 (DNB) 和荷兰金融市场监管局 (AFM) 监管, 并于荷兰商会以摩根大通有限责任公司分行的名义注册登记, 其注册编号为 72610220。在**丹麦**, 本文件是由**摩根大通有限责任公司哥本哈根分行** (即德国摩根大通有限责任公司联属公司) 分派, 其注册办事处位于 Kalvebod Brygge 39-41, 1560 København V, Denmark, 已获德国联邦金融监管局 (BaFin) 授权, 并由 BaFin、德国中央银行和欧洲中央银行共同监管。摩根大通有限责任公司哥本哈根分行 (即德国摩根大通有限责任公司联属公司) 同时须受丹麦金融监管局 (Finanstilsynet) 监管, 并于丹麦金融监管局以摩根大通有限责任公司分行的名义注册登记, 编号为 29010。在**瑞典**, 本文件由**摩根大通有限责任公司斯德哥尔摩分行**分派, 其注册办事处位于 Hamngatan 15, Stockholm, 11147, Sweden, 已获德国联邦金融监管局 (BaFin) 授权, 并由 BaFin、德国中央银行和欧洲中央银行共同监管。摩根大通有限责任公司哥本哈根分行同时须受瑞典金融监管局 (Finansinspektionen) 监管, 并于瑞典金融监管局以摩根大通有限责任公司分行的名义注册登记。在**比利时**, 本文件由**摩根大通有限责任公司——布鲁塞尔分行**分派, 其注册办事处位于 35 Boulevard du Régent, 1000, Brussels, Belgium, 已获德国联邦金融监管局 (BaFin) 授权, 并由 BaFin、德国中央银行和欧洲中央银行共同监管。摩根大通有限责任公司布鲁塞尔分行同时须受比利时国家银行 (NBB) 及比利时金融服务及市场管理局 (FSMA) 监管, 并已于比利时国家银行行政注册处登记注册, 注册编号为 0715.622.844。在**希腊**, 本文件由**摩根大通有限责任公司——雅典分行**分派, 其注册办事处位于 3 Haritos Street, Athens, 10675, Greece, 已获德国联邦金融监管局 (BaFin) 授权, 并由 BaFin、德国中央银行和欧洲中央银行共同监管。摩根大通有限责任公司雅典分行同时须受希腊银行监管, 并已于希腊银行行政注册处以摩根大通有限责任公司分行的名义登记注册, 注册编号为 124。雅典商会注册号为 158683760001; 增值税注册号为 99676577。在**法国**, 本文件由**摩根大通有限责任公司巴黎分行**分派, 其注册办事处位于 14, Place Vendôme 75001 Paris, France, 已获德国联邦金融监管局 (BaFin) 授权, 并由 BaFin、德国中央银行和欧洲中央银行共同监管, 注册编号为 842 422 972, 摩根大通有限责任公司巴黎分行亦受法国银行业监察委员会(Autorité de Contrôle Prudentiel et de Résolution (ACPR))及法国金融市场监管局 (Autorité des Marchés Financiers(AMF)) 监管。在**瑞士**, 本文件由 **J.P. Morgan (Suisse) S.A.** 分派, 其注册办事处位于 rue du Rhône, 35, 1204, Geneva, Switzerland, 作为瑞士一家银行及证券交易商, 在瑞士由瑞士金融监督管理局 (FINMA) 授权并受其监管。

在**香港**, 本文件由**摩根大通银行香港分行**分派, 摩根大通银行香港分行受香港金融管理局及香港证监会监管。在香港, 若您提出要求, 我们将会在不收取您任何费用的情况下停止使用您的个人资料以作我们的营销用途。在**新加坡**, 本文件由**摩根大通银行新加坡分行**分派。摩根大通银行新加坡分行受新加坡金融管理局监管。交易及咨询服务及全权委托投资管理服务由摩根大通银行香港分行/新加坡分行向您提供 (提供服务时会通知您)。银行及托管服务由摩根大通银行香港分行/新加坡分行向您提供 (提供服务时会通知您)。本文件的内容未经香港或新加坡或任何其他法律管辖区的任何监管机构审阅。建议您审慎对待本文件。假如您对本文件的内容有任何疑问, 请必须寻求独立的专业人士意见。对于构成《证券及期货法》及《财务顾问法》项下产品广告的材料而言, 本营销广告未经新加坡金融管理局审阅。摩根大通银行 (JPMorgan Chase Bank, N.A.) 是依据美国法律特许成立的全国性银行组织; 作为一家法人实体, 其股东责任有限。

关于**拉美**国家, 本文件的分派可能会在特定法律管辖区受到限制。

在**澳大利亚**, 由**摩根大通银行 (ABN 43 074 112 011/AFS牌照号码: 238367) 和摩根大通证券(ARBN 109293610) 发行。**

「摩根大通」是指摩根大通及其全球附属公司和联属公司。「摩根大通私人银行」是摩根大通从事私人银行业务的品牌名称。本文件仅供您个人使用, 未经摩根大通的允许不得分发给任何其他人士, 且任何其他人士均不得使用, 分派或复制本文件的内容供作非个人用途。如您有任何疑问或不欲收取这些通讯或任何其他营销资料, 请与您的摩根大通团队联络。