英伟达（NVDA.O）

百川终将归海，AI 奇点到来

全球人工智能算力中枢，AI 需求推动强劲增长。硬件端，英伟达目前形成了“GPU+CPU+DPU”的产品组合，成为纵横数据中心、游戏显卡、专业可视化、自动驾驶等多个赛道的算力之王。在数据中心赛道，2023 年的 AI 芯片市场中英伟达出货量约占 60-70%。在游戏显卡赛道，英伟达占据了消费级独立显卡的 80%以上出货量。软件端，我们认为 DGX Cloud 长期有望成为英伟达数据中心业务的第二增长曲线。

AI 算力需求暴增带来英伟达收入利润表现强劲。FY2024Q3 英伟达收入为 181.20 美元、yoy+206%，其中数据中心、游戏、专业可视化、自动驾驶分别为 145/29/4/3 亿美元，yoy+279%/+81%/+108%/4%。2024Q3 财季公司 non-GAAP 净利约 100 亿美金、non-GAAP 净利润率达 55%。

需求：AI 算力需求可以延续多久？英伟达数据中心业绩的可持续性，来自于 AI 算力需求的可持续性。1）训练端，更多国家、企业将入场 AI 军备战争，模型的参数数据量也更大。2）推理端，端侧 AI 的逐步落地、AI 应用向更多科技和制造领域破圈，均带来更强的推理算力需求。

我们粗略测算：1）训练端：基于假设，至 2030 年全球累计需要相当于 2000 万张 H100 的等量算力需求。2）推理：基于假设，至 2030 年全球累计需要相当于超 1.16 亿张 A30 的等量算力需求。

供给：龙头面对搅局者。AI 芯片供给的竞争方还有：1）AMD 和 Intel 等数据中心 GPU 新手。2）谷歌 TPU、微软 Athena 等云厂商自研芯片。

英伟达对于来自竞争对手的挑战，亦做了充分的准备：1）软硬件产品上，公司在硬件产品上持续迭代新品，在软件架构上持续延续优势。英伟达望在 2024 年发布 Hopper 架构 H200、还有望提前发布其下一代 GPU Blackwell B100。CUDA 架构开发者和下载量亦在持续提升。2）上下游生态上，英伟达一方面通过投资参股等方式绑定下游企业的算力需求，一方面通过上百亿美金采购承诺额锁定上游产能。

基于 CoWoS 的产能增长、对英伟达不同产品线的产能分配等假设，按英伟达财年维度，我们测算：2025/2026 财年，H100 的出货量望达 209 万张/155 万张、H200 望达 35 万张/62 万张、B100 望达 23 万张/143 万张。

投资建议：首次覆盖给予“买入”评级。我们预计 2024-2026 财年公司收入将达 594/986/1282 亿美元，同比增长 120%/66%/30%。Non-GAAP 净利润 313/524/610 亿美元，同比增长 274%/68%/16%。考虑到英伟达净利润高速增长，我们认为英伟达合理市值为 20964 亿美元、对应股价为 840.6 美金，对应 40x FY2025e P/E（FY 2025 财年为 2024 年 1 月至 2025 年 1 月），首次覆盖给予“买入”评级。

风险提示：下游 AI 应用不及预期、数据中心算力芯片竞争超预期、AI 行业政策监管超预期、假设和测算误差风险。

财务指标	FY2022A	FY2023A	FY2024E	FY2025E	FY2026E
营业收入（百万美元）	26,914	26,974	59,368	98,607	128,150
增长率 yoy（%）	61	0	120	66	30
Non-GAAP 净利（百万美元）	11,259	8,366	31,281	52,411	60,991
增长率 yoy（%）	79.4	-25.7	273.9	67.5	16.4
EPS 最新摊薄（美元/股）	4.44	3.34	12.54	21.01	24.46
净资产收益率（%）	36.6	19.8	56.2	48.9	36.3
P/E（倍）	162.4	216.1	57.5	34.3	29.5
P/S（倍）	66.8	66.7	30.3	18.2	14.0

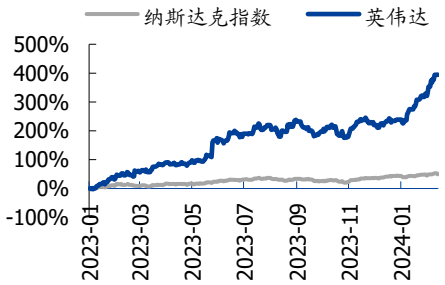
资料来源：公司公告，国盛证券研究所　注：股价为 2 月 13 日收盘价

买入（首次）

股票信息

行业	海外
2 月 13 日 收盘价(美元)	721.28
总市值(百万美元)	1,780,119
总股本(百万股)	2,468
其中自由流通股(%)	100%

股价走势



作者

分析师　夏君
执业证书编号：S0680519100004
邮箱：xiajun@gszq.com

分析师　朱若菲
执业证书编号：S0680522030003
邮箱：zhuruofei@gszq.com

相关研究

财务报表和主要财务比率

资产负债表（百万美元）

会计年度	FY2022A	FY2023A	FY2024E	FY2025E	FY2026E
流动资产	28829	23073	53572	107223	170214
现金及现金等价物	1990	3389	26709	68732	119921
有价证券	19218	9907	9907	9907	9907
应收账款	4650	3827	8423	13990	18182
存货	2605	5159	7399	12653	19253
其他流动资产	366	791	1134	1940	2952
非流动资产	15358	18109	18283	18391	18458
固定资产	2778	3807	3981	4089	4156
商誉	4349	4372	4372	4372	4372
无形资产	2339	1676	1676	1676	1676
其他非流动资产	5892	8254	8254	8254	8254
资产总计	44187	41182	71855	125613	188672
流动负债	4335	6563	8870	14281	21078
短期债务	0	1250	1250	1250	1250
应付账款	1783	1193	1711	2926	4452
其他流动负债	2552	4120	5909	10105	15376
非流动负债	13240	12518	12518	12518	12518
长期债务	10946	9703	9703	9703	9703
其他非流动负债	2294	2815	2815	2815	2815
负债合计	17575	19081	21388	26799	33596
少数股东权益	0	0	0	0	0
普通股	3	2	2	2	2
资本公积	10385	11971	11971	11971	11971
留存收益	16235	10171	38538	86885	143146
其他综合损益	-11	-43	-43	-43	-43
归属母公司股东权益	26612	22101	50468	98815	155076
负债和股东权益	44187	41182	71855	125613	188672

现金流量表（百万美元）

会计年度	FY2022A	FY2023A	FY2024E	FY2025E	FY2026E
经营活动现金流	9108	5641	25320	44023	53188
净利润	9752	4368	28367	48347	56262
折旧摊销	1174	1544	1826	1892	1933
营运资金变动	-3363	-2207	-4873	-6216	-5006
其他经营现金流	1545	1936	0	0	0
投资活动现金流	-9830	7375	-2000	-2000	-2000
资本支出	-976	-1833	-2000	-2000	-2000
其他投资现金流	-8854	9208	0	0	0
筹资活动现金流	1865	-11617	0	0	0
借款所得	0	0	0	0	0
股份回购	-	-10039	0	0	0
分红	-399	-398	0	0	0
其他筹资现金流	2264	-1180	0	0	0
现金净增加额	1143	1399	23320	42023	51188

利润表（百万美元）

会计年度	FY2022A	FY2023A	FY2024E	FY2025E	FY2026E
营业收入	26914	26974	59368	98607	128150
营业成本	9439	11618	16662	28495	43358
研发费用	5268	7339	8880	14223	19223
销售及行政费用	2166	2440	2725	3749	4873
收购终止成本	0	1353	0	0	0
营业利润	10041	4224	31101	52139	60697
其他损益	-100	-43	389	136	136
利润总额	9941	4181	31490	52275	60833
所得税费用	189	-187	3123	3928	4571
净利润	9752	4368	28367	48347	56262
EPS（美元/股）	4	2	11	19	23
Non-GAAP 净利润	11259	8366	31281	52411	60991
Non-GAAP EPS（美元/股）	4	3	13	21	24

主要财务比率

会计年度	FY2022A	FY2023A	FY2024E	FY2025E	FY2026E
成长能力					
营业收入（%）	61.4	0.2	120.1	66.1	30.0
营业利润（%）	37.3	15.7	52.4	52.9	47.4
归属母公司净利润（%）	125.1	-55.2	549.4	70.4	16.4
获利能力					
毛利率（%）	64.9	56.9	71.9	71.1	66.2
净利率（%）	36.2	16.2	47.8	49.0	43.9
ROE（%）	36.6	19.8	56.2	48.9	36.3
ROIC（%）	37.0	18.9	54.2	48.2	35.9
偿债能力					
资产负债率（%）	39.8	46.3	29.8	21.3	17.8
净负债比率（%）	7.5	9.7	50.4	68.3	76.5
流动比率	6.7	3.5	6.0	7.5	8.1
速动比率	5.6	2.9	5.1	6.5	7.2
营运能力					
总资产周转率	0.7	0.6	1.1	1.0	0.8
应收账款周转率	7.6	6.4	9.7	8.8	8.0
应付账款周转率	18.0	18.1	40.9	42.5	34.7
每股指标（元）					
Non-GAAP EPS（最新摊薄）	4.44	3.34	12.54	21.01	24.46
每股经营现金流（最新摊薄）	3.65	2.26	10.15	17.65	21.33
每股净资产（最新摊薄）	10.67	8.86	20.24	39.62	62.18
估值比率					
P/E	162.4	216.1	57.5	34.3	29.5
P/B	67.6	81.4	35.6	18.2	11.6
P/S	66.8	66.7	30.3	18.2	14.0

资料来源：公司公告，国盛证券研究所　注：股价为 2 月 13 日 收盘价

内容目录

图表目录

1. 全球领先的算力平台

1.1 全球算力之源

英伟达（NVIDIA）由黄仁勋、Chris Malachowsky 和 Curtis Priem 创立于 1993 年。

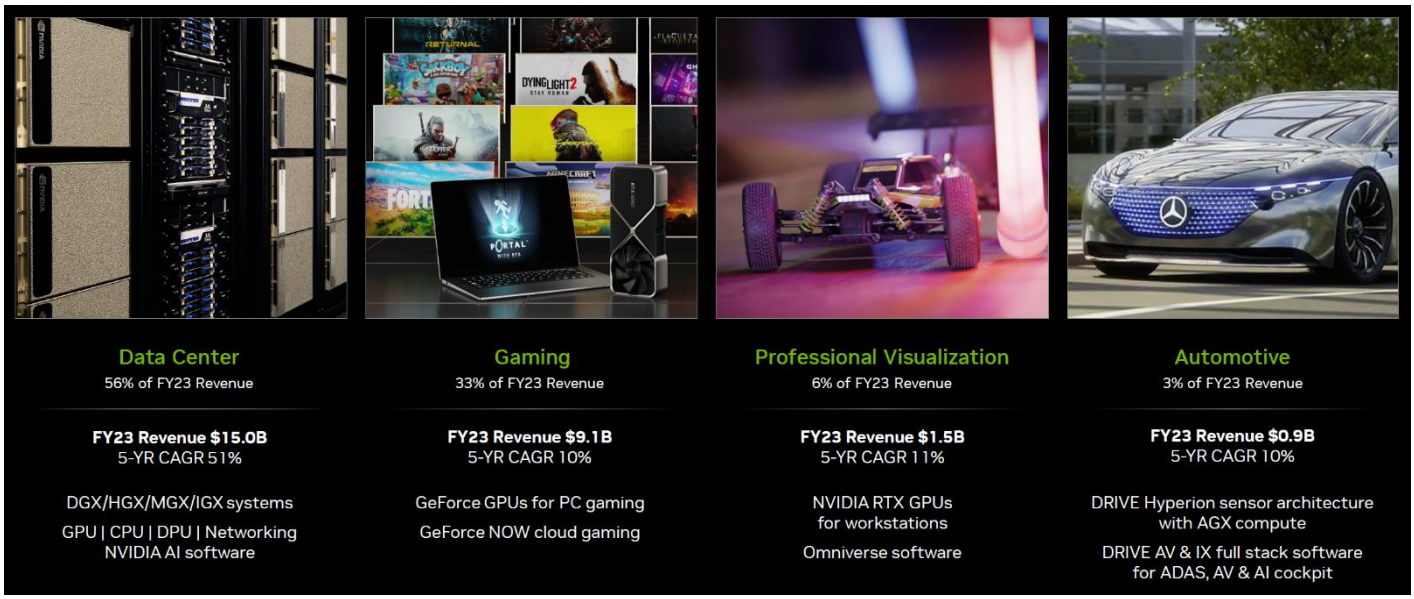
1999 年，英伟达推出 Geforce 256，被称为 GPU（Graphics Processing Unit）的定义者。起初的若干年，英伟达核心产品是游戏显卡——这一阶段的战役在经历了与关键对手 ATI 的缠斗、与重要客户微软和索尼的诉讼和合作、与两大 CPU 巨头 Intel 和 AMD 的合纵连横之后，终于以在 2006 年以 AMD 收购 ATI、2009 年 Intel 暂时取消自研 GPU 计划为标志而暂落下帷幕。

此后的时间里，一方面，英伟达将芯片产品扩展至更多行业赛道——如 2008 年苹果的 Macbook 搭载英伟达 GeForce 9400MG、2012 年特斯拉的 Model S 搭载英伟达自动驾驶芯片、2019-2021 年加密货币浪潮中的 GTX1060 和 CMP 系列；另一方面英伟达也在积极向 DPU 和 CPU 环节延展——2020 年英伟达收购 Mellanox Technologies 从而将芯片产品扩展至 DPU，2021 年英伟达在 GTC 2021 大会推出了基于 ARM 架构的首款 CPU 并命名为 Grace。

至此，英伟达形成了“GPU+CPU+DPU”的产品组合，成为横贯数据中心、游戏显卡、专业可视化、自动驾驶等多个赛道的算力之王。

1.1.1 业务一览：全球算力之王

图表 1：NVIDIA 核心业务赛道构成



资料来源：NVIDIA 财报 PPT，国盛证券研究所

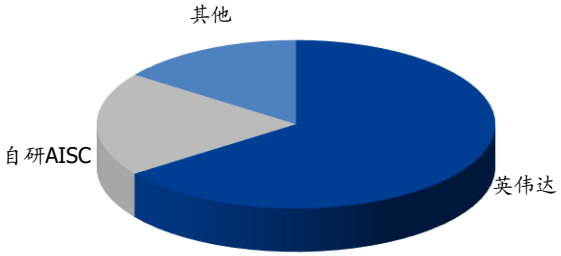
英伟达的算力芯片产品遍及数据中心、游戏显卡、专业可视化、自动驾驶等多个行业赛道。公司作为业内的算力之王，其统治力从相应赛道市占率可见一斑：

- 在数据中心赛道，Trendforce 数据显示，2023 年的 AI 芯片市场中英伟达出货量约占 60-70%，几家互联网巨头的自研 ASIC 芯片约占 20%。

  当然，如果仅看数据中心 GPU 产品，则英伟达 A100、H100 等产品在模型训练等方面几乎没有可替代的对手选项。

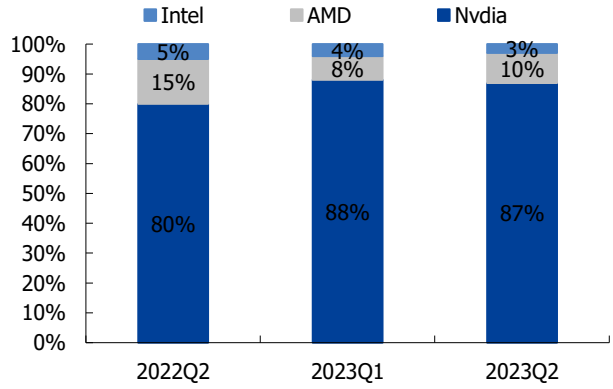
- 在游戏显卡赛道，JPR 数据显示，英伟达占据了 PC 独立显卡的 80%以上出货量。收购了 ATI 的 AMD 当前则在 10%左右的市占率浮动。

图表 2：英伟达市场地位——2023AI 芯片出货量：英伟达占 60-70%



资料来源：TrendForce，国盛证券研究所

图表 3：英伟达市场地位——PC 独立 GPU 出货量：英伟达超过 80%



资料来源：JPR，国盛证券研究所

图表 4：英伟达核心股东情况

股东名称	持股比例
Vanguard 集团	8.25%
贝莱德	7.31%
FMR	5.22%
道富集团	3.59%
黄仁勋	3.51%
普徕仕	2.26%
Geode Capital Management	1.96%
摩根大通	1.41%
摩根士丹利	1.32%
挪威中央银行	1.08%
北美信托	1.07%

资料来源：Bloomberg，截止 2024/2/8，国盛证券研究所

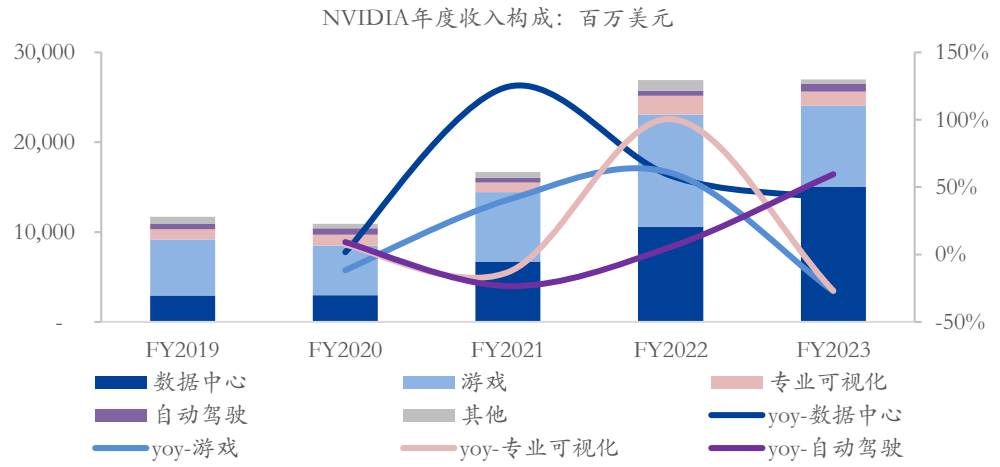
1.1.2 财务构成：AI 需求推动数据中心业务强劲增长

英伟达的核心芯片产品线包括数据中心、游戏、专业可视化、自动驾驶等。

- 截至 2023 财年(结束于 2023 年 1 月)，英伟达年度收入约 270 亿美金，同比持平。其中，数据中心业务占比 56%，游戏业务占比 34%，专业可视化占比 6%，自动驾驶业务占比 3%，其他业务占比 2%。
- 截至 2024Q3 财季（结束于 2023 年 10 月），英伟达季度收入约 180 亿美金，同比增长 206%。其中，数据中心业务占比 80%，游戏业务占比 16%，专业可视化占比 2%，自动驾驶业务占比 1%。

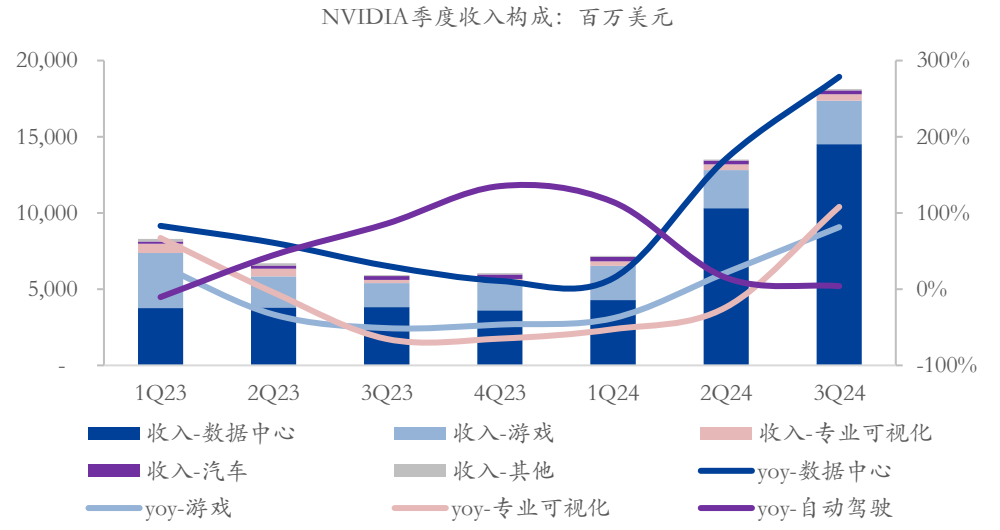
  2024Q3 财季，得益于 AI 算力需求强劲，英伟达数据中心业务同比增长 279%，单业务收入亦创新高。

图表 5：NVIDIA 收入构成：年度



资料来源：公司公告，国盛证券研究所

图表 6：NVIDIA 收入构成：季度

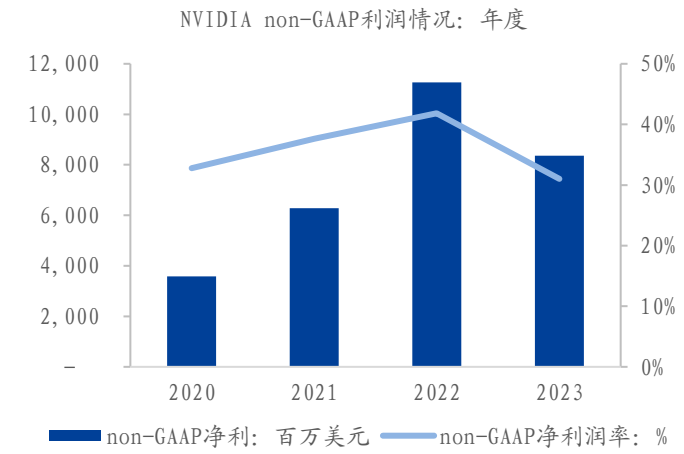


资料来源：公司公告，国盛证券研究所

同样，得益于 AI 算力需求暴增带来的数据中心 GPU 供不应求，英伟达利润表现也非常强劲：

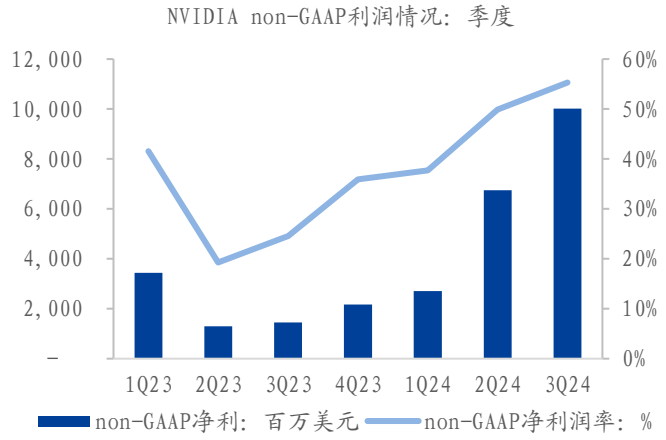
- 截至 2023 财年（结束于 2023 年 1 月），英伟达年度 GAAP 口径利润约 43.7 亿美金，non-GAAP 口径利润约 83.7 亿美金。公司 non-GAAP 净利润率达 31%。
- 截至 2024Q3 财季（结束于 2023 年 10 月），英伟达季度 GAAP 口径利润约 92.4 亿美金，non-GAAP 口径利润约 100.2 亿美金。公司 non-GAAP 净利润率达 55%。

图表 7：NVIDIA 利润情况：年度



资料来源：公司公告，国盛证券研究所

图表 8：NVIDIA 利润情况：季度



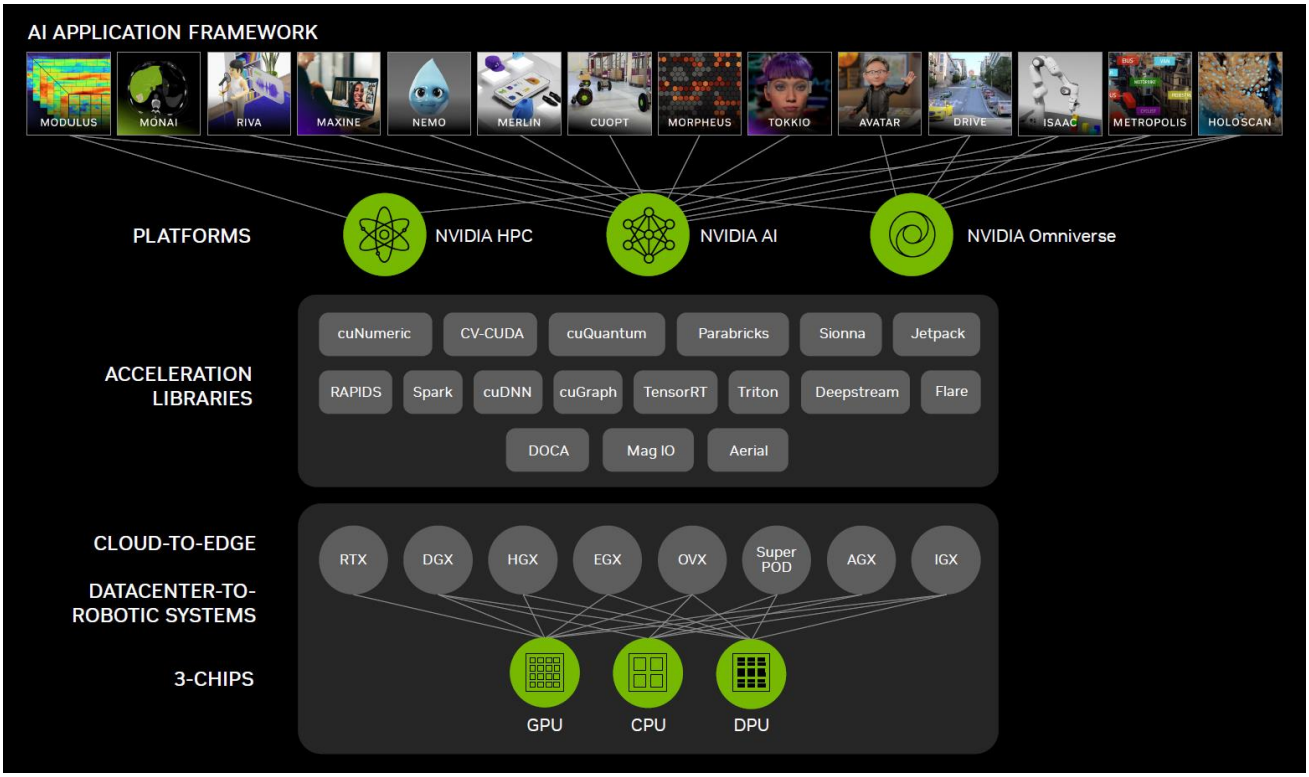
资料来源：公司公告，国盛证券研究所

1.2 硬件：“GPU+CPU+DPU”，纵横多个行业赛道

1）数据中心

英伟达的数据中心业务涵盖自下而上、从硬件产品到软件平台的全栈产品。其中硬件部分包含 GPU、CPU、DPU 三大类别芯片；软件方面包括 CUDA 并行编程模型、CUDA- x 应用程序加速库、应用程序编程接口、或 API、SDK 和工具、以及特定领域的应用程序框架等；平台端则包含 NVDIA HPC、NVDIA AI、NVDIA Omniverse 等平台。英伟达计算平台专注于在超大规模、云、企业、公共部门和边缘数据中心加速最计算密集型的工作负载，如人工智能、数据分析、图形和科学计算。

图表 9：NVIDIA 数据中心业务：自下而上，从硬件产品到软件平台



资料来源：NVIDIA 财报 PPT，国盛证券研究所

数据中心 GPU 是英伟达的王牌产品，公司主要产品包括训练/推理芯片 A100、H100、L40、L40S 等，以及推理芯片 A10、A30 等。2024 年公司将推出性能更强的 H200、B100 等。

图表 10：NVIDIA 数据中心产品：GPU

GPU	A100 SXM	H100 SXM	L40	L40S	A10	A30
架构	Ampere	Hopper	Ada Lovelace	Ada Lovelace	Ampere	Ampere
市场定位	训练/推理	训练/推理	训练/推理	训练/推理	推理	推理
Tensor FP16 峰值算力：TFLOPS	624	1979	181.05	362.05	125	165
Tensor TF32 峰值算力：TFLOPS	312	989	90.5	183	62.5	82
GPU 显存	HBM2e	HBM3	支持 ECC 的 48GB GDDR6	支持 ECC 的 48GB GDDR6	GDDR6	HBM2
显存容量	80GB	80GB	48GB	48GB	24GB	24GB
显存带宽	2039 GB/s	3.35 TB/s	864 GB/s	864 GB/s	600 GB/s	933 GB/s
互连技术	NVLink：600 GB/s PCIe 4.0: 64 GB/s	NVLink：900GB/s PCIe 5.0：128GB/s	PCIe 4.0: 64 GB/s	PCIe 4.0: 64 GB/s	PCIe 4.0: 65 GB/s	NVLink：200 GB/s PCIe 4.0:64 GB/s

资料来源：公司官网，国盛证券研究所

除了 GPU 之外，英伟达也扩大其他数据中心处理器产品组合：

数据中心 CPU 方面，英伟达也在加速布局，比如推出数据中心 CPU 产品 NVIDIA Grace CPU 超级芯片。此外，英伟达也推出适用于大规模 AI 和 HPC 应用的突破性加速 CPU——NVIDIA Grace Hopper 超级芯片。

图表 11：英伟达数据中心产品：CPU

	Grace CPU	Grace Hopper
GPU HBM capacity (GB)		96GB HBM3 144GB HBM3e
Grace CPU cores (number)		Up to 72 cores
GPU HBM bandwidth (GB/s)		4TB/s HBM3 4.9TB/s HBM3e
浮点算力	FP64：peak 7.1 TFLOPS	FP64 ：34 teraFLOPS。FP32： 67 teraFLOPS
显存带宽	Up to 1TB/s	Up to 512GB/s （grace） 900 GB/s bidirectional（hopper）
互连技术	NVLink-C2C bandwidth：900GB/s PCIe links：Up to 8x PCIe Gen5 x16 option to bifurcate	NVLink-C2C bandwidth: 900 GB/s bidirectional PCIe links：Up to 4x PCIe x16 (Gen5)
核心控制 Core count	144 Arm Neoverse V2 Cores with 4x128b SVE2	72 Arm Neoverse V2 cores
低功耗内存 LPDDR5X size	240GB, 480GB and 960GB	Up to 480GB

资料来源：公司官网，国盛证券研究所

DPU 方面，NVIDIA BlueField 网络平台为全球数据中心提供动力，凭借强大的计算能力以及用于网络、存储和安全加速的内置软件定义硬件加速器，BlueField 可为各种环境中的多种工作负载提供安全的加速基础设施。DPU 产品包含 NVIDIA BlueField-3、BlueField-2、BlueField-3 SuperNIC 等。

图表 12：英伟达数据中心产品：DPU

	BlueField-2	BlueFhield-3	NVIDIA BlueField-3 SuperNIC
网络接口	以太网-10/25/50/100Gb/s 的双端口，或 200Gb/s 的单端口 InfiniBand-EDR/HDR100(100Gb/s)双端口或 HDR(200Gb/s)的单端口	1 或 2 端口高达 400Gb/s 以太网或 NDR InfiniBand 网络连接	1 或 2 端口高达 400Gb/s 以太网或 NDR InfiniBand 网络连接
PCI Express 接口	8 或 16 通道的 PCIe Gen 4.0 PCIe 交换机（含多达 8 个下行端口）	32 通道第五代 PCIe 多达 16 个下行端口的 PCIe 交换拆分	32 通道第五代 PCIe 多达 16 个下行端口的 PCIe 交换拆分
ARM 核心	多达 8 个 Armv8A72 核心（64 位）流水线	多达 16 个 ARMv8.2+A78 Hercules 核心（64 位）	多达 16 个 ARMv8.2+A78 Hercules 核心（64 位）
DDR4 DIMM 支持	单个 DDR4DRAM 控制器 8GB/16GB 板载 DDR4	双 DDR5 5600MT/s DRAM 控制器 32GB 板载 DDR5 内存	双 DDR5 5600MT/s DRAM 控制器 32GB 板载 DDR5 内存

资料来源：公司官网，国盛证券研究所

2）游戏显卡

英伟达针对游戏市场的产品包括用于游戏台式机和笔记本电脑的 GeForce RTX 和

GeForce GTX GPU，以及用于玩 PC 游戏的 GeForce NOW 云游戏平台，用于电视高质量流媒体的 SHIELD、以及用于游戏机的系统芯片(SoC)和开发服务。在 2023 财年，英伟达推出了基于 Ada Lovelace 架构的 GeForce RTX 40 系列游戏 GPU。

图表 13：产品梳理：英伟达消费级 GPU 主要产品

GeForce 型号	NVIDIA CUDA® 核心数量（个）	加速频率 (GHz)	基础频率 (GHz)	标准显存配置	显存位宽
RTX 4090 D	14592	2.52	2.28	24 GB GDDR6X	384 位
RTX 4080 SUPER	10240	2.55	2.29	16 GB GDDR6X	256 位
RTX 4080	9728	2.51	2.21	16 GB GDDR6X	256 位
RTX 4070 Ti SUPER	8448	2.61	2.34	16 GB GDDR6X	256 位
RTX 4070 Ti	7680	2.61	2.31	12 GB GDDR6X	192 位
RTX 4060 Ti	4352	2.54	2.31	16 GB GDDR6 或 8 GB GDDR6	128 位
RTX 3090 Ti	10752	0.86	1.56	24 GB GDDR6X	384 位
RTX 2080 Ti	4352	1.64	1.35	11 GB GDDR6	352 位
GTX 1660 Ti	1536	1770	1500	6GB GDDR6	192 位

资料来源：公司官网，国盛证券研究所

图表 14：竞品对比：AMD 消费级 GPU 主要产品

Radeon 型号	计算单元	基准频率	加速频率	游戏频率	峰值半精度 (FP16) 性能	峰值单精度 (FP32) 性能	最大显存	显存类型	显存位宽	最大显存带宽
RX 7900 XTX	96	-	最高可达 2500 MHz	2300 MHz	123 TFLOPS	61 TFLOPS	24GB	GDDR6	-	最高可达 960 GB/s
RX 7900M	72	-	-	1825 MHz	77.05 TFLOPS	38.52 TFLOPS	16GB	GDDR6	-	最高可达 576 GB/s
RX 7700S	32	-	-	2200 MHz	41 TFLOPS	20.5 TFLOPS	8GB	GDDR6	128-bit	最高可达 288 GB/s
RX 6950 XT	80	1890 MHz	最高可达 2310 MHz	2100 MHz	47.31 TFLOPS	23.65 TFLOPS	16GB	GDDR6	256-bit	最高可达 576 GB/s
RX 6850M XT	40	-	-	2463 MHz	26.6 TFLOPS	13.3 TFLOPS	12GB	GDDR6	192-bit	最高可达 432 GB/s
RX 6750 XT	40	2150 MHz	最高可达 2600 MHz	2495 MHz	26.62 TFLOPS	13.31 TFLOPS	12GB	GDDR6	192-bit	最高可达 432 GB/s
RX 6750 GRE 12GB	40	2321 MHz	最高可达 2581 MHz	2439 MHz	26.43 TFLOPS	13.21 TFLOPS	12GB	GDDR6	192-bit	最高可达 384 GB/s
RX 6650 XT	32	2055 MHz	最高可达 2635 MHz	2410 MHz	21.59 TFLOPS	10.79 TFLOPS	8GB	GDDR6	128-bit	最高可达 280 GB/s
RX 6550M	16	-	-	2560 MHz	11.6 TFLOPS	5.8 TFLOPS	4GB	GDDR6	64-bit	最高可达 144 GB/s
RX 6550S	16	-	-	2170 MHz	9.9 TFLOPS	4.9 TFLOPS	4GB	GDDR6	64-bit	最高可达 128 GB/s

资料来源：AMD 官网，国盛证券研究所

3） 专业可视化

英伟达专业可视化产品的适用范围包括设计和制造以及数字内容创建。例如设计和制造包括计算机辅助设计、建筑设计、消费产品制造、医疗仪器和航空航天。数字内容创作包括专业视频编辑和后期制作、电影特效和广播电视图形。主要硬件产品包括 Ada Lovelace 架构的专业卡 RTX 6000 等、Ampere 架构的 RTX A6000 系列、Turing 架构的 T1000 等。

图表 15：英伟达专业可视化主要产品

产品	GPU Memory	Display Ports	Max Power Consumption	Form Factor	Thermal
NVIDIA Ada Lovelace Architecture					
RTX 6000	48GB GDDR6 with ECC	4x DisplayPort 1.4a*	300W	4.4” (H) x 10.5” (L) dual slot	Active
RTX 5000	32GB GDDR6 with ECC	4x DisplayPort 1.4a*	250W	4.4” (H) x 10.5” (L) dual slot	Active
RTX 4500	24GB GDDR6 with ECC	4x DisplayPort 1.4a	210W	4.4” (H) x 10.5” (L) dual slot	Active
NVIDIA Ampere Architecture					
A800 40GB Active	40GB HBM2	Not equipped	240W	4.4” (H) x 10.5” (L) dual slot	Active
RTX A6000	48GB GDDR6 with ECC	4x DisplayPort 1.4a*	300W	4.4” (H) x 10.5” (L) dual slot	Active
RTX A5500	24GB GDDR6 with ECC	4x DisplayPort 1.4a*	230W	4.4" (H) x 10.5" (L) dual slot	Active
NVIDIA Turing Architecture					
T1000	4GB | 8GB GDDR6	4x Mini DisplayPort 1.4	50W	2.713" (H) x 6.137" (L) single slot	Active
T600	4GB GDDR6	4x Mini DisplayPort 1.4	40W	2.713" (H) x 6.137" (L) single slot	Active
T400	2GB | 4GB GDDR6	3x Mini DisplayPort 1.4	30W	2.713" (H) x 6.137" (L) single slot	Active

资料来源：公司官网，国盛证券研究所

4）自动驾驶

NVIDIA 的汽车业务由自动驾驶、智能座舱、电动汽车计算平台和信息娱乐平台解决方案组成，将以 DRIVE Hyperion 品牌为自动驾驶市场提供完整的端到端解决方案。硬件方面，英伟达自动驾驶芯片主要包含 Xavier、Orin、Thor 等。

图表 16：英伟达当前自动驾驶芯片

产品	Xavier	Orin	Thor
发布时间	2018	2019	2022
量产时间	2020	2022	2025E
制程	12nm	7nm	/
功耗	30W	45W	/
算力	30 TOPS	254 TOPS	2000 TOPS

资料来源：公司公告、维科网、玩车教授、IT 之家、电子工程世界，国盛证券研究所

1.3 软件及平台：云服务望成长为第二曲线

当然，英伟达作为全球领先的算力平台，在硬件产品之外，亦为客户提供了多维度的软件平台服务，包括但不限于：

- DGX Cloud：云服务平台，可提供 NVIDIA DGX AI 超级计算专用集群，并配以 NVIDIA AI 软件。DGX Cloud 不仅包括算力，还包括一整套“AI 训练即服务”解决方案。
- Omniverse：元宇宙应用平台，使用 OpenUSD 开发工业元宇宙应用，适用于汽车、建筑、工程、施工和运营、媒体和娱乐，以及制造行业等。
- GeForce Now：云游戏平台，支持玩家绑定 Steam、Epic Games 账号，通过 NVIDIA GeForce Now 云游戏来体验已有游戏库中的游戏。
- Automobile Drive：自动驾驶平台，其中开放式 NVIDIA DRIVE® SDK 为开发者提供了自动驾驶所需的所有构建块和算法堆栈，该软件有助于开发者更高效地构建和部署各种先进的自动驾驶应用程序。

其中，DGX Cloud 作为英伟达数据中心业务在算力芯片产品之外的重要业务方向，将数据中心业务扩展到了算力和模型训练等相关的云服务方面。NVIDIA AI 则包括加速计算、基础 AI 软件、预训练模型和“AI 代工厂”。预训练模型和“AI 代工厂”包括语言模型 NEMO、视觉模型 PICASSO、生物学模型 BIONEMO、游戏模型 NVIDIA ACE、生成式 AI 模型（包括 GPT、T5 和 Llama 等）等等。

我们认为，DGX Cloud 有望将算力和模型训练相关业务以更易得的方式提供给企业客户，长期有望成为英伟达数据中心业务的第二增长曲线。

2. 需求：AI 算力需求可以延续多久

2.1 AI 需求：对下一个时代的押注，谁也不能松懈

2022 年，OpenAI 推出 ChatGPT，带来了人工智能浪潮。此后，全球互联网及云服务大厂陆续加入大模型的军备战争，AI 算力需求快速提升。英伟达数据中心业绩的可持续性，来自于算力需求的可持续性。

2.1.1 训练端：谁在边际增加 AI 算力投入？

人工智能实力的提升，是一个互联网及云服务企业甚至于一个国家都不能错过的战斗。当前我们看到，AI 军备战争已经从 2023 年的少数几家互联网及云服务大厂，向更多地区的更多企业和部门扩展。接下来，更多国家和企业将入场 AI 军备战争：

- 更多国家入场：法国、英国、德国、瑞典、越南、新加坡、印度、日本等国家和地区开始加大 AI 投入。
- 更多企业增加投入：Meta、OpenAI、以及微软、谷歌等均在加大 AI 投入。
- 模型更大：随着多模态的发展、各家模型厂商之间的竞争加剧，模型的参数数据量也更大。

图表 17：AI 训练需求增加：更多国家开始加入 AI 军备竞赛

国家	时间	事件
越南	2024 年 1 月	越南通信传媒部近日发布了一项计划，到 2025 年，越南至少拥有一个越南语大语言模型。越南企业 Vingroup 旗下子公司 VinBigData 开发，第一个向公众开放的越南生成式人工智能大模型 ViGPT。
新加坡	2024 年 1 月	新加坡上个月宣布了一项计划，将针对印尼语、马来语和泰语研发大语言模型。
日本	2024 年 1 月	日本政府联合日本电气公司、富士通、软银等大型科技公司投入数亿美元，开发日语大型语言模型。
英国	2023 年 11 月	英国政府宣布将投资 2.25 亿英镑研发人工智能（AI）超级计算机“Isambard-AI”，助推英国成为 AI 领域的全球领导者。
德国	2023 年 10 月	德国的 Julich 超级计算中心也宣布了其建设下一代人工智能超级计算机的计划，使用接近 24000 个 Grace Hopper Superchips 和 Quantum-2 InfiniBand，将其提升为全球最强大的人工智能超级计算机，拥有超过 90 exaflops 的人工智能性能。
印度	2023 年 10 月	印度电子和信息技术部于 2023 年 10 月 14 日发布了《印度人工智能 2023 计划》。重点关注了包括计算机基础设施建设、人工智能研究和创新能力提升、国家机器人战略草案草拟、人工智能芯片开发、印度数据集建设等问题。
	2024 年 1 月	印度数据中心运营商 Yotta 计划向美国智能芯片制造商英伟达追加购买价值 5 亿美元的 AI 芯片，用以强化其人工智能云服务能力。2023 年 12 月，印度数据中心运营商 Yotta 订购约 16000 颗英伟达 H100 芯片，预计到 2024 年 1 月，将有 4096 个 GPU 投入 Yotta 的人工智能云服务。
法国	2023 年 6 月	巴黎初创公司 Mistral AI 宣布获得一轮超过 1 亿美元的种子资金，以构建类似 ChatGPT 的大型语言模型和生成式 AI。该公司预计将于明年推出首个 AI 产品，目前的估值已经超过了 2 亿欧元。
瑞典	2023 年 12 月	AI Sweden 与 RISE 和 WASP WARA Media & Language 一起，为北欧语言（主要是瑞典语）开发了一个大规模的生成语言模型。

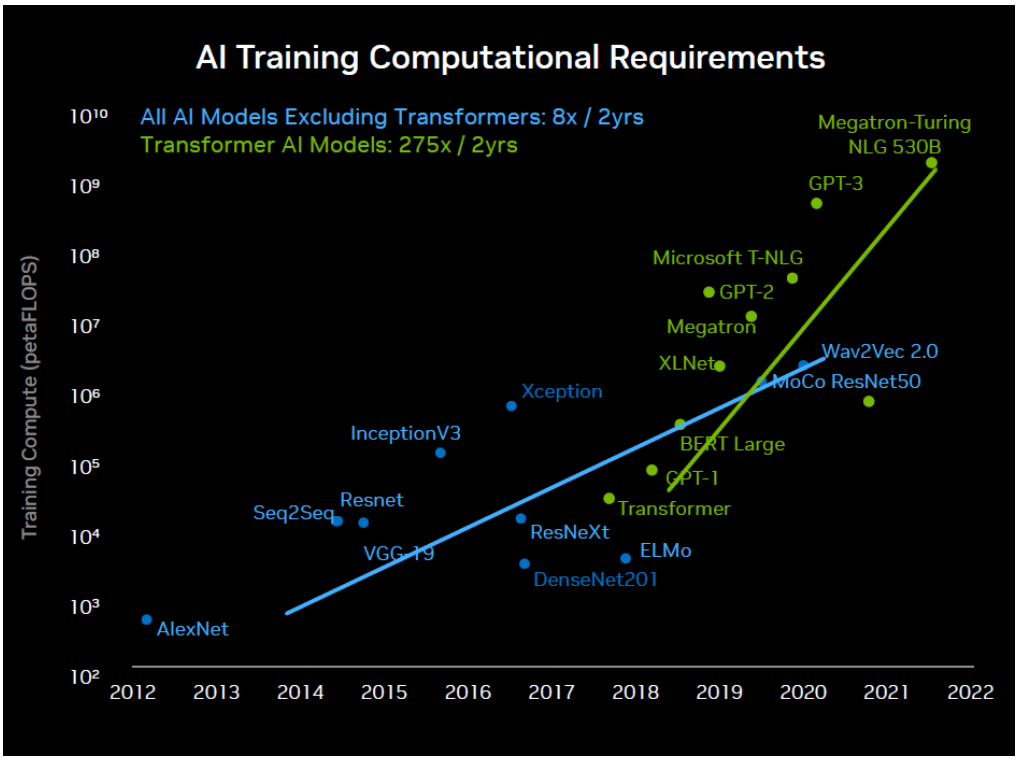
资料来源：英伟达业绩会、环球网、澎湃新闻、第一财经、AIse 官网、印度政府电子与信息技术部、腾讯网，国盛证券研究所

图表 18：AI 训练需求增加：更多企业开始加入 AI 军备竞赛

公司	时间	事件
Meta	2024 年 1 月	Meta 首席执行官扎克伯格宣布，到 2024 年底前 Meta 将购买约 35 万张英伟达 H100，包括其他 GPU 将有大约 60 万 H100 的等效算力。
Amazon	2024 年 2 月	为了加强其云业务，亚马逊正在向聊天机器人制造商 Anthropic 投资高达 40 亿美元。亚马逊首席财务官布莱恩·奥尔萨夫斯基（Brian Olsavsky）在电话会议上表示，亚马逊预计今年的资本支出将增加，以支持 AWS 的增长，包括对生成式 AI 和大型语言模型的额外投资。
谷歌	2024 年 2 月	2024 年资本支出将继续增加，以支持 AI 的持续投资。
微软	2024 年 2 月	1）微软首席财务官艾米·胡德（Amy Hood）预计，微软的资本开支将在 2024 年一季度显著环比扩大。2）微软管理层判断，随着未来推理需求增长，云资源的消耗速度还将进一步提升。
OpenAI	2024 年 1 月	OpenAI 计划投资千亿美元自建芯片工厂，以应对全球 AI 芯片需求激增。
苹果	2024 年 2 月	库克在财报会上指出，苹果正全力以赴地投入大量时间和精力于 AI 的研发，并计划在今年晚些时候公布更多细节。

资料来源：澎湃新闻、财联社、亿邦动力、财经十一人、搜狐网、中国经济周刊，国盛证券研究所

图表 19：AI 训练需求增加：AI 模型加速迭代，参数量大幅提升



资料来源：公司公告，国盛证券研究所

2.1.2 推理端：哪些 AI 场景和应用在增加？

我们看到，AI 推理相关的算力需求正在海量袭来，而背后的驱动力则包括端侧 AI 的逐步落地、AI 应用从文娱内容领域向更多科技和制造领域扩展等方面。

- 端侧 AI 落地

  近期 AI 大模型功能在硬件端落地的浪潮开启：AI PC、AI 手机、AI+可穿戴新型便携产品等迭起，AI 赋能硬件产品更智能、交互更顺畅、提升用户体验。

  ✓ 2023 年 11 月，Humane 发布无屏幕可穿戴设备 AI Pin，背后是 OpenAI 的 GPT-4 为其提供 AI 能力，可实现语音通话、写文稿、听音乐、处理电子邮件、实时翻译等任务，未来计划增加导航和购物功能。

  ✓ 2024 年 1 月，联想携 40 多款产品亮相 CES 2024，其中包括十余款 AI PC。联想宣布个人 AI 助理——Lenovo AI Now 将在今年上半年部署到产品上市。

  ✓ 2024 年 1 月，三星发布首款 AI 手机 Galaxy S24，全面集成了三星自研的前沿生成式 AI 模型 Gauss，同时，谷歌 AI 大模型 Gemini nano 在其中得到全面应用，为搜索、通话、短信、相机等都配置了 AI 功能。

  我们预期，端侧 AI 产品的快速普及将为 AI 推理带来大量需求。

图表 20：AI 推理需求增加：云到端

终端应用	案例公司	相关产品	功能
AI+可穿戴:Pin	Humane	AI Pin	AI Pin 重不足 40 克，可吸附在衣服等物体表面，无需实体屏幕，可语交互音/投影在手掌交互。AI 模型加持下，AI Pin 可实现写文稿、听音乐、实时翻译等任务，未来计划增加导航和购物功能。
AI+可穿戴:吊坠	Rewind	Pendant	产品形态类似项链吊坠，可以捕捉所说和所听的内容，转录、加密并完全存储在手机上，提供个性化 AI 能力。
AI+可穿戴:眼镜	Meta	Ray-Ban	拍摄照片、60 秒视频以及听音乐、接电话等，且支持 Meta AI。
AI+可穿戴:AR/VR	Meta	Meta Quest 3	或接入 Meta AI，可以实现对话，提供做饭、旅游、写作的建议等。
AI+可穿戴:MR	苹果	Vision Pro	可以接入多模态 AI 助手 Otter，以视频为输入，能完成多模态感知、推理、和上下文学习。
AI+可穿戴:耳机	讯飞	iFLYBUDS Nano	现场录、通话录、音视频录。搭载生成式 AI 会议助理 VIAIM，智能提炼、总结关键信息，快速生成会议摘要、提取待办事项、并对待办事项进行跟进。
AI+手机	谷歌	Pixel 8 系列	更好的拍照和视频功能，翻译、实时转录消息，检测并过滤垃圾电话，检测用户是否遭遇严重车祸并呼叫紧急服务，Fitbit 将使用生成式 AI 为用户带来个性化指导、动态锻炼建议。引入 Bard 后，Google Assistant 将个性化功能与大模型的推理和生成能力相结合，实现听、说、影响处理能力的全面升级。
AI+手机	三星	Galaxy S24	自研生成式 AI 产品 Gauss 将面向 AI 聊天、AI 代码、AI 图片等领域，同时，谷歌 AI 大模型 Gemini nano 在其中得到全面应用。
AI+手机	苹果	尚未推出	探索 AI 嵌入应用程序。
AI+手机	华为	mate 60	华为手机智慧助手小艺具备 AI 大模型能力，在交互、生产力提升和个性化服务三个方向上增强。
AI+手机	小米	小米 14 系列	搭载骁龙 8gen3 芯片,基于小米自研的 AI 大模型，澎湃 OS 将实现小爱输入助手、WPS 随手拍、AI 妙画、AI 搜图、AI 写真、AI 扩图、实时字幕等功能。
AI+手机	Vivo	X100 系列	Vivo OriginOS 4 正式亮相、将大模型能力与系统结合，X100 系列手机将全球首发搭载天玑 9300 旗舰芯片。
AI+手机	OPPO	Find X6 系列	OPPO 正式推出 AndesGPT 并接入新操作系统 ColorOS 14，Find X6 系列等 6 款机型将首发升级正式版。小布助手支持了内容创作、用机助手、智能摘要、智能消除等各类 AIGC 能力。
AI+手机	荣耀	Magic6 系列	荣耀 Magic6 系列将支持自研 70 亿端侧 AI 大模型。
AI+PC	联想	联想 AIPC	可以创建本地知识库和运行个人基础大模型，还支持 AI 计算和自然交互。除此之外，联想还通过大模型压缩技术，保证了个人隐私和数据安全，使得 AIPC 能够在本地运行个人大模型，不需要依赖云端操作。
AI+PC	三星	Galaxy Book 4	发布首款人工智能（AI）笔记本 Galaxy Book 4 系列笔记本电脑，预计该系列将于 2024 年 1 月在韩国上市，之后推向其他地区。
AI+智能音箱	亚马逊	Alexa	接入为语音交互定制的大模型，可提供更自然的对话功能和智能家居控制。
AI+智能音箱	小米	小爱同学	小米团队结合大模型的对话特点升级了小爱的交互模型。小爱拥有优秀的上下文理解能力；全新沉浸式的对话形态，生成式的结果。具体而言，小爱同学可以写周报、做旅游攻略、制定健身计划，甚至写代码等。
AI+智能车	特斯拉	FSD V12	FSD V12 打造基于神经网络的端到端大模型。
AI+智能车	蔚来	智驾	基于 BEV+Tansformer 开展无图拓城计划。
AI+智能车	小鹏	智驾	基于 BEV+Tansformer 开展无图拓城计划。
AI+智能车	理想	座舱+智驾	理想发布自研认知大模型“Mind GPT”，包括对话生成、语言理解、知识问答、逻辑推理等在内的各项能力变得更安全，更准确，也更有逻辑。
AI+智能车	华为	问界 M9	问界 M9 在智能座舱和智能驾驶双“天花板”基础上，搭载的黑科技包括：鸿蒙智能座舱、华为智能驾驶、HUAWEI xPixel、HUAWEI AR-HUD、AI 大模型、传感联邦、HUAWEI SOUND 等。
AI+机器人	特斯拉	Optimus	特斯拉的全自动驾驶系统 FSD 直接被应用在 Optimus 身上，机器人采用了与汽车一样的视觉感知，使用摄像头输入数据，以神经网络进行计算。
AI+机器人	英伟达	Eureka（机器人训练工具）	Eureka 可以教会机器人复杂的运动控制技能，比如转笔、打开抽屉和柜子、抛球和接球、操作剪刀。
AI+机器人	Meta	Habitat3.0(机器人训练工具)	Habitat 3.0 是第一个支持在多样化、逼真的室内环境中，就人机交互任务进行大规模训练的模拟器。

资料来源：各公司官网、官方微博、澎湃新闻、财联社、IT 之家、华尔街见闻、量子位、金融界、钛媒体、新浪、中国证券网、经济观察网、36 氪、快科技、每日经济新闻、科创版日报，国盛证券研究所

- 领域破圈

我们认为，接下来生成式 AI 的应用，除了可以在内容领域以外，会在更多的领域和圈层落地。

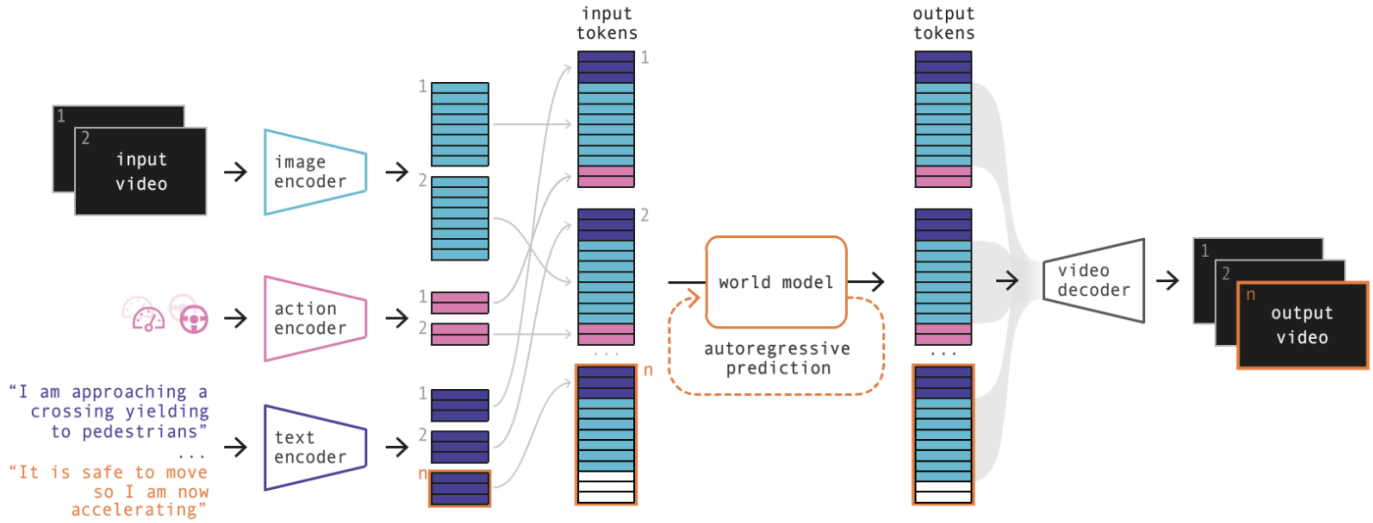
✓ 首先，以自动驾驶领域为例，各类 AI 工具被广泛地应用在数据合成、4D 标注、感知模型、决策规划模型、以及当前的端到端模型探索中。其中，由于 1）数据采集成本日益提高、2）真实场景的数据采集涉及隐私安全信息，3）有效 corner case 的收集密度太低等原因，自动驾驶的训练往往面临数据不足的问题。基于此，自动驾驶领域一些企业，如 Wayve，已经开始通过生成式 AI 模型来创建驾驶场景视频，用以更好地辅助自动驾驶端侧模型的开发。

英国创业公司 Wayve 在 2023 年 6 月首次推出了 GAIA-1（Generative Artifitial Intelligengce for Autonomy）、并在 2023 年 9 月更新了最新进展。GAIA-1 模型核心是一个基于自回归 Transformer 的世界模型（world model）：在输入视频、文本、动作指引后，能预测序列中下一组图像 token；这些预测的图像 token 不仅在视觉上连贯、而且和此前的文字和动作指引保持一致。随后，视频解码器（video diffusion model）将这些图像 token 转换回像素空间。

除了 Wayve 以外，Tesla 也在尝试通过建立仿真场景来辅助自动驾驶模型的训练。Tesla 在 CVPR 2023 workshop 展示了其“General World Model”，市场普遍认为其除了可以为自动驾驶决策规划模型的训练提供“模拟器”环境外、后续作用还可能体现在自动驾驶算法本身。

我们认为，生成式 AI 模型有望大大降低自动驾驶模型训练的门槛、以及提升决策规划的能力天花板。

图表 21：自动驾驶场景仿真：GAIA-1 模型框架



资料来源：《Scaling GAIA-1: 9-billion parameter generative world model for autonomous driving》，作者为 Anthony Hu、Lloyd Russell 等，国盛证券研究所

✓ 其次，在生物及材料科学领域，我们也看到了生成式 AI 在蛋白质预测、新型材料生成等“AI For Science（AI4S）”方面的巨大潜力。

谷歌 Deepmind 旗下的 Alphafold 是生物医学领域比较早出圈的 AI 工具，此前就可以进行单链蛋白质的预测、以及后续扩展至具有多条蛋白质链的复合物。2023 年 10

月底，新一代 Alphafold 进一步加强，不仅可以预测蛋白质结构，还可以进行对核酸、小分子配体等生物分子结构的预测。该工具有助于加速生物医学的进展。

谷歌 Deepmind 旗下的另一个工具 GNoMe，则是将类似能力应用在了新材料的发现上。GNoMe 基于图神经网络对晶体材料进行预测和筛选。当前 GNoMe 发现了 220 万种新晶体材料，而且将预测材料稳定性的准确率从 50%拉高到 80%。

微软 MatterGen 的突破则在于，可以针对所需要的特性，直接生成相应的新型材料。MatterGen 基于类似 Diffusion Model 的方法，为晶体材料选取了定制的扩散过程、得出基础模型。然后引入适配器模块，在带有属性标签的附加数据集上对基础模型进行微调，最终引导生成的结果符合目标属性约束。这一技术有望大大加快设计所需特性材料的速度。

图表 22：AI 推理需求增加：AI 应用向更多基础研究领域扩展

模型及工具名称	所属公司	时间	技术	成果
新一代 Alphafold	谷歌 Deepmind	2023 年 10 月	原有 Alphafold 是单链蛋白预测的突破，Alphafold-Multitimer 扩展到具有多条蛋白质链的复合物。新一代 Alphafold 在此基础上进一步拓展。	新一代 Alphafold 不仅可以预测蛋白质结构，还可以进行对核酸、小分子配体等生物分子结构的预测。该工具有助于加速生物医学的进展。
GNoMe	谷歌 Deepmind	2023 年 11 月	GNoMe 基于图神经网络（GNN），将已知的稳定材料生成候选结构，然后对这些候选结构进行筛选。筛选得出的结果进行结构稳定性的验证，随后作为新的训练数据再给到 Gnome、用以改进预测能力。	谷歌使用 GNoMe 工具，发现了 220 万种新晶体材料，而且将预测材料稳定性的准确率从 50%拉高到 80%。
MatterGen	微软	2023 年 12 月	基于类似 Diffusion Model 的方法，为晶体材料选取了定制的扩散过程、得出基础模型。然后引入适配器模块，在带有属性标签的附加数据集上对基础模型进行微调，最终引导生成的结果符合目标属性约束。	MatterGen 可以针对所需要的特性，直接生成相应的新型材料。这些生成的材料具有结构的独特性和新颖性。

资料来源：新智元、量子位、澎湃新闻，国盛证券研究所

从内容生成，到自动驾驶场景仿真、到材料定制，我们认为后续生成式 AI 在科技、制造等研究及生产领域可以带来更多推理需求、也创造更多产业价值。

2.2 定量测算：模型训练与推理，全球需要多少卡

我们用粗略的测算，来估计当前全球企业在 AI 模型的训练和推理过程中所需要的算力芯片的量级：

首先从训练的角度，我们以 GPT-4（根据 SemiAnalysis，约 1.8 万亿参数、13 万亿训练数据）作为基础，假设后续几年全球各国大模型数量持续增加、模型参数继续攀升，则按我们的测算，至 2030 年，全球累计需要相当于 2000 万张 H100 的等量算力需求。

图表 23：训练所需 GPU 需求–按 H100 测算

	2023e	2024e	2025e	2026e	2027e	2028e	2029e	2030e
国家数	3	5	5	6	7	8	9	10
科技巨头数/国家	3	5	5	5	5	5	5	5
模型数/科技巨头企业	3	4	5	5	5	5	5	5
模型参数扩容速度：		10%	10%	10%	10%	10%	10%	10%
Token 数：十亿	13000	14300	15730	17303	19033	20937	23030	25333
模型参数量：十亿	1800	1980	2178	2396	2635	2899	3189	3508
Flops/token/模型参数量-训练	6	6	6	6	6	6	6	6
峰值算力 TFLOPS-H100 SXM	989	989	989	989	989	989	989	989
算力利用率假设	21.3%	21.3%	21.3%	21.3%	21.3%	21.3%	21.3%	21.3%
总计卡数需求：百万张	0.6	2.6	3.9	5.6	7.9	11.0	14.9	20.1
每年新增卡数需求：百万张		2.0	1.3	1.7	2.3	3.0	4.0	5.1

资料来源：《The economics of large language models》、作者为 SUNYAN，semianalysis，国盛证券研究所测算

其次从推理的角度，我们同样以 GPT-4（根据 SemiAnalysis，约 1.8 万亿参数、13 万亿训练数据）作为基础，假设后续几年全球各国大模型数量持续增加、模型参数继续攀升、应用迭代带来的用户访问用量持续提升，则按我们的测算，至 2030 年，全球累计需要超过 1.16 亿张相当于 A30 的等量算力需求。

图表 24：推理所需 GPU 需求-按 A30 测算

	2023e	2024e	2025e	2026e	2027e	2028e	2029e	2030e
每次查询次数/访问	8	10	11	13	14	16	17	19
每次查询输入字数	50	50	50	50	50	50	50	50
每次查询输出字数	200	200	200	200	200	200	200	200
字数/token	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
token 数/访问	2667	3167	3667	4167	4667	5167	5667	6167
国家数	3	5	5	6	7	8	9	10
科技巨头数/国家	3	5	5	5	5	5	5	5
模型数/科技巨头企业	3	4	5	5	5	5	5	5
单个模型每月访问：亿次	5	6	6	7	8	9	9	10
单模型访问次数/年：百万	6000	6840	7680	8520	9360	10200	11040	11880
模型参数扩容速度		10%	10%	10%	10%	10%	10%	10%
模型参数量：十亿	1800	1980	2178	2396	2635	2899	3189	3508
Flops/token/模型参数量-推理	2	2	2	2	2	2	2	2
峰值算力 TFLOPS – A30	165	165	165	165	165	165	165	165
算力利用率假设	21.3%	21.3%	21.3%	21.3%	21.3%	21.3%	21.3%	21.3%
总计卡数需求：百万张	1.4	7.7	13.8	23.0	36.4	55.1	81.0	115.9

资料来源：《The economics of large language models》、作者为 SUNYAN，semianalysis，国盛证券研究所测算

3. 供给：龙头面对搅局者

3.1 AI 芯片江湖：扶持 AMD、发力自研芯片

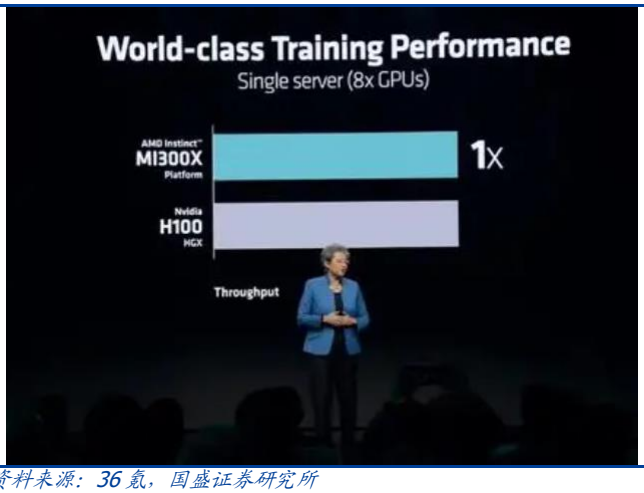
在算力芯片如此紧缺的当下，众多互联网及云服务厂商当然也不能把鸡蛋放在一个篮子里——既不够安全、又太贵。当前的 AI 芯片赛道，除了种子选手英伟达之外，还有两类重要的阵营：

- 以 AMD 和 Intel 为代表的 GPU 专业级新选手，
- 以谷歌 TPU、微软 Athena 等为代表的云厂商自研芯片。

以 AMD 为例，AMD 于 2023 年 6 月发布 AMD Instinct MI300X GPU 和 AMD Instinct MI300A APU。在硬件性能角度，MI300X 堪与 H100 一战：

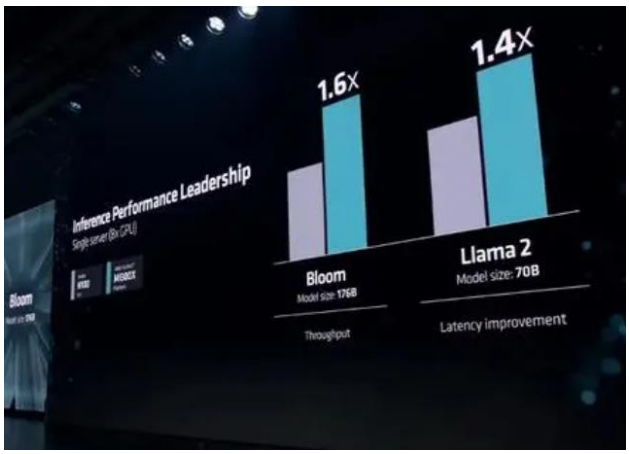
- 内存：内存容量较 H100 的 80GB 提高 2.4 倍至 192GB，内存带宽从 3.4TB/s 提升 1.6 倍至 5.3TB/s。
- HPC 表现：算力性能高达 H100 的 2.4 倍。
- AI 表现：算力性能高达 H100 的 1.3 倍。

图表 25：训练性能对比：AMD MI300X vs 英伟达 H100



资料来源：36 氪，国盛证券研究所

图表 26：推理性能对比：AMD MI300X vs 英伟达 H100



资料来源：36 氪，国盛证券研究所

根据 AMD 的表述，就 AMD Instinct MI300X platform 与英伟达 H100 HGX 的比较而言，在训练方面二者性能相当，在推理方面则 MI300X platform 的推理速度是 H100 HGX 的 1.4 倍（Llama2）到 1.6 倍（Bloom）。

图表 27：供给端竞争：AMD MI300X vs 英伟达相关 GPU 参数比较

对比	MI300X	A100 SXM	H100 SXM	H200 SXM	B100
GPU 架构	AMD CDNA3	NVIDIA Ampere	NVIDIA Hopper	NVIDIA Hopper	Blackwell
技术制程	TSMC 5nm | 6nm FinFET	TSMC 7nm	TSMC 4nm	TSMC 4nm	TSMC 3nm
FP64	81.7 TFLOPS	9.7 TFLOPS	34 TFLOPS	34 TFLOPS	-
FP32	163.4 TFLOPS	19.5 TFLOPS	67 TFLOPS	67 TFLOPS	-
TF32	1307.4 TFLOPS	312 TFLOPS	989 TFLOPS	989 TFLOPS	-
FP16	2614.9 TFLOPS	624 TFLOPS	1979 TFLOPS	1979 TFLOPS	-
BF16	2614.9 TFLOPS	624 TFLOPS	1979 TFLOPS	1979 TFLOPS	-
FP8	5229.8 TFLOPS	-	3958 TFLOPS	3958 TFLOPS	-
INT8	5229.9 TFLOPS	1248 TFLOPS	3958 TOPS	3958 TOPS	-
GPU Memory	192 GB	80GB	80 GB	141 GB	192GB
GPU Memory Type	HBM3	HBM2e	HBM3	HBM3e	HBM3e
GPU Memory Bandwidth	5.3 TB/s	2039 GB/s	3.3 TB/s	4.8 TB/s	-
Interconnect	Infinity Fabric:896 GB/s	NVLink:600 GB/s	NVLink:900 GB/s	NVLink:900 GB/s	-
TDP	750 W	400 W	700 W	700 W	-
晶体管数量	1530 亿	540 亿	800 亿	800 亿	1780 亿

资料来源：英伟达官网、AMD 官网、量子位、快科技、机器之心 Pro、澎湃新闻、芯智讯、IT 之家、科创板日报、tweaktown、每日经济新闻，国盛证券研究所

同样，很多对 AI 芯片需求较高的大厂亦早早开始布局自研芯片，如谷歌 TPU、微软 Athena、亚马逊 Tranium 等。

图表 28：国际巨头 AI 芯片布局

公司	芯片	发布时间	代际	制程	设计	用途
亚马逊	Trainium	2020	1	5nm	自研	训练
亚马逊	Inferentia2	2022	2	5nm	自研	推理
谷歌	TPU v5	2023	5	5nm/7nm	自研	训练
谷歌	Maple	2025（E）	1	5nm	marvell technology	
谷歌	Cypress	2025（E）	1	5nm	自研	
微软	Azure Maia 100	2023	1	5nm	自研	训练和推理
微软	Azure Cobalt 100	2023	1	5nm	自研	
微软	Athena	2024（E）	1	5nm	自研	训练和推理
微软	Cascade	2024（E）	1	5nm	自研	
特斯拉	D1	2021	1	7nm	自研	训练
Meta	MTIA v1	2023	1	7nm	自研	推理

资料来源：The Information reporting、亚马逊官网、热点科技、新智元、钛媒体、半导体行业观察、芯东西、The Information、机器之心、快科技、澎湃新闻，国盛证券研究所

其中以谷歌为例，谷歌自 2015 年发布 TPU v1 以来，不断迭代升级，在 TPU v2 时已经可以支持训练。其在 2021 年 Q2 发布的 TPU v4 通过光互连实现可重配置和高可拓展性，采用 7nm 工艺，峰值算力达 275TFLOPS，性能大幅提升。根据谷歌发布的论文《TPU v4: An Optically Reconfigurable Supercomputer for Machine Learning with Hardware Support for Embeddings》，使用 TPU v4 芯片进行嵌入训练时，相比于使用 TPU v3 芯片，可以获得 2.7 倍的性能提升。

图表 29：云服务厂商发力自研芯片：以 Gooogle TPU 为例

	TPU v1	TPU v2	TPU v3	TPU v4i	TPU v4
发布时间	2015Q2	2017Q3	2018Q4	2020Q1	2021Q2
DNN target	推理	训练和推理	训练和推理	推理	训练和推理
Peak TFLOPS/Chip	92 (8b int)	46 (bf16)	123 (bf16)	138 (bf16/8b int)	275 (bf16/8b int)
TDP(Watts)Chip/System 散热设计功耗（W）	75/220	280/460	450/660	175/275	-
Chip Technology	28nm	16nm	16nm	7nm	7nm

资料来源：《Ten Lessons From Three Generations Shaped Google's TPUv4i》、作者为 Norman P. Jouppi、Doe Hyun Yoon 等，《TPU v4: An Optically Reconfigurable Supercomputer for Machine Learning with Hardware Support for Embeddings》、作者为 Norman P. Jouppi、George Kurian 等，芯智讯，国盛证券研究所

3.2 英伟达的破局

我们认为，英伟达对于来自竞争对手的挑战，亦具备充分的信心、以及做了其充分的准备：

- 软硬件产品上，公司在硬件产品上持续迭代新品，在软件架构上持续延续优势。

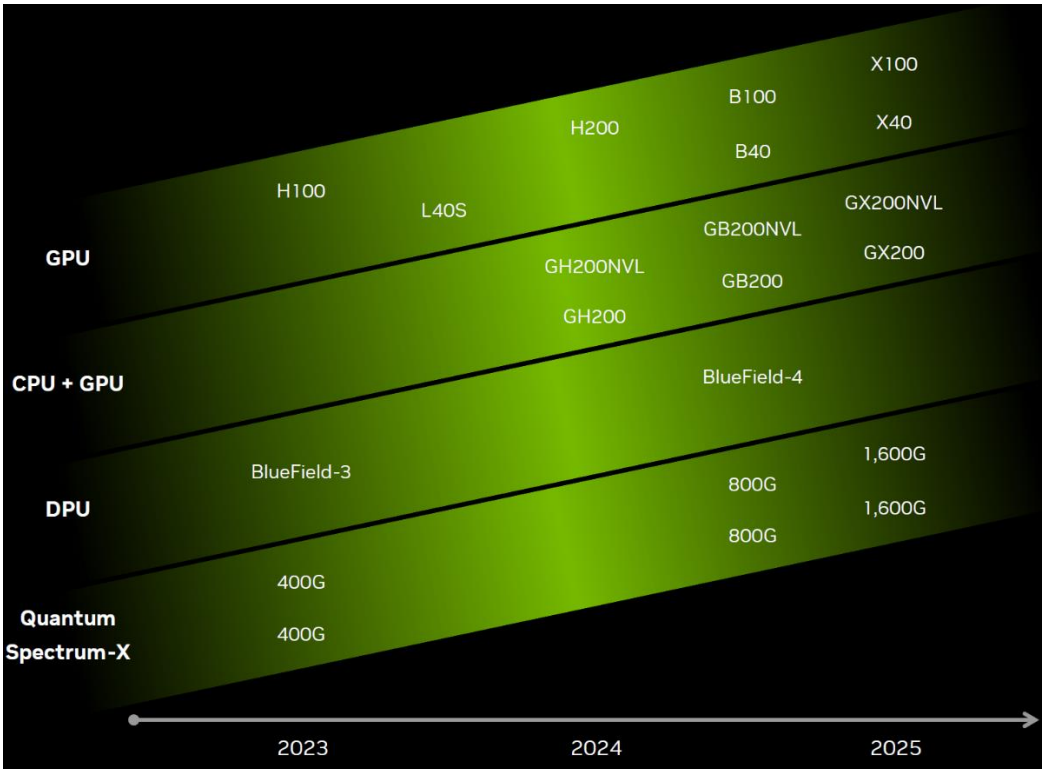
  英伟达望在 2024 年发布 Hopper 架构 H200、还有望提前发布其下一代 GPU Blackwell B100。CUDA 架构开发者和下载量亦在持续提升。

- 上下游生态上，英伟达一方面通过投资参股等方式绑定下游企业的算力需求，一方面通过上百亿美金采购承诺协议锁定上游产能。

3.2.1 软硬件产品：加速迭代下一代硬件产品、CUDA 持续保持优势

在硬件方面，英伟达或于 2024 年推出 Hopper 架构 H200、Blackwell 架构 B100。面对 AMD Instinct MI300 系列的汹汹来势，英伟达或提前其 B100 产品的推出和交付以做应对。据英伟达，2024 年推出的 Blackwell 架构 B100 GPU，在 GPT-3 175B 推理性能标竿方面击败 A100、H100 及 H200，其 AI 表现性能将是 Hopper 架构 H200 GPU 两倍以上。

图表 30：硬件产品：NVIDIA 数据中心产品 pipeline



资料来源：NVIDIA 财报 PPT，国盛证券研究所

在软件架构方面，AMD 为了更好地兼容 CUDA 平台，ROCm 复制了 CUDA 的技术栈，支持 HIP（类 CUDA）和 OpenCL 两种 GPU 编程模型，开发者可以用类似 CUDA 的方式为 AMD 的 GPU 产品编程，从而在源代码层面兼容 CUDA。

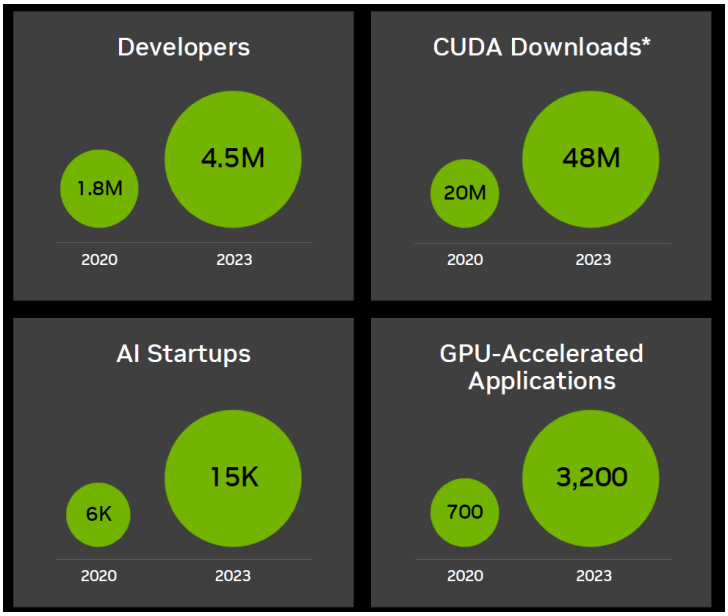
图表 31：软件生态对比：英伟达 vs. AMD

	Nvidia CUDA	AMD ROCm
厂商	英伟达	AMD
推出时间	2006 年	2016 年
编程	CUDA 使英伟达的 GPU 能够执行使用 C、C++、Fortran、OpenCL、DirectCompute 和其他语言编写的程序。	支持 HIP（类 CUDA）和 OpenCL 两种 GPU 编程模型。其中 HIP 的编程语法与 CUDA 相似，开发者可以用类似 CUDA 的方式为 AMD 的 GPU 产品编程，从而在源代码层面上兼容 CUDA。
排他性	仅适用于英伟达硬件。	ROCm 支持多种加速器厂商和架构，提供了开放的可移植性和互操作性。

资料来源：芯世相、快科技、AMD 官网、IT 之家，国盛证券研究所

但在实操角度，英伟达 CUDA 架构具有较明显的先发优势。CUDA 架构当前拥有 450 万开发者，2023 年软件下载量达 4800 万次，15000 家创业企业使用 CUDA 架构。网络效应驱动 CUDA 架构受众持续增加。

图表 32：NVIDIA 软件 CUDA 架构优势



资料来源：NVIDIA 财报 PPT，国盛证券研究所

3.2.2 上下游生态：绑定下游、锁定上游

除了产品端过硬之外，英伟达在需求端对下游企业的生态绑定、在供给端对上游供应商的产能锁定，也使得英伟达有着更加稳定的上下游关系。

- **下游：大举投资 AI 模型企业**

  从 2023 年到 2024 年，英伟达投资了大量大模型及 AI 相关企业。通过投资这些企业，英伟达进一步扩张了 AI 版图、我们认为有助于其绑定下游潜在需求。

图表 33：英伟达投资 AI 企业

投资公司	时间	投资方	融资金额	AI 业务
Adept AI	2023 年 3 月	英伟达、微软等	3.5 亿美元	Adept 的旗舰基础模型 ACT-1 与现有的生成式人工智能工具不同，因为它能够解释用户对软件工具的高级自然语言请求，并直接为它们执行任务。
CoreWeave	2023 年 4 月	英伟达、Friedman、Gross	2.2 亿美元	CoreWeave 提供了对云端 Nvidia GPU 的十几个 SKU 的访问权限，包括 H100s、A100s、A40s 和 RTXA6000s，用于人工智能和机器学习、视觉效果和渲染、批处理和像素流等用例。
Cohere	2023 年 5 月	英伟达、甲骨文等	2.7 亿美元	Cohere 的生成式 AI 模型主要面向的是企业级客户，包括全球流媒体平台、服装公司以及使用该平台简化客户服务或提高内容审核能力的公司。
Runway	2023 年 6 月	谷歌、英伟达	1.41 亿美元	Runway 利用计算机图形学和机器学习方面的最新进展发布了 Gen1 和 Gen2 两代视频生成模型，其中 Gene1 还需要提供原源频，而 Gen2 仅需要几个单词就能生成短视频。
Inflection AI	2023 年 7 月	英伟达、微软和谷歌前首席执行官埃里克·施密特	13 亿美元	Inflection AI 大部分资金将用于增强计算能力，以开发更强大的基础模型，其最新的 AI 基础模型名为 Inflection-2。新模型的训练速度更快、成本更低，但仍然可以处理大量运算（1025 FLOP）。
AI21 Labs	2023 年 8 月	英伟达、三星、谷歌	1.55 亿美元	AI21 Labs 为企业提供基于文本的生成人工智能服务，技术包括领先的大型语言模型和神经符号技术。
Hugging Face	2023 年 8 月	英伟达、谷歌、亚马逊 、 Salesforce, 、AMD、英特尔、IBM 和高通等	2.35 亿美元	Hugging Face 打造了一个平台，人工智能开发人员可以在其中共享代码、模型、数据集，并使用该公司的开发工具让开源人工智能模型更轻松地运行。
Imbue	2023 年 9 月	英伟达	2 亿美元	与 ChatGPT 这样的大规模人工智能基础模型不同，Imbue 瞄准的是 AI 代理：一种可以模拟人类决策来完成复杂任务的计算系统，Imbue 专注与创建专为推理而定制的基础模型。
Twelve Labs	2023 年 10 月	英伟达	970 万美元	Twelve Labs 开发了能够理解视频内容的超大规模人工智能模型
Mistral AI	2023 年 12 月	英伟达、Salesforce 等	4.5 亿欧元	Mistral AI 专注于聊天机器人和生成人工智能工具的开源软件。几个月前在开源 Apache 2.0 许可下发布了 Mistral 7B，这是其第一个大型语言模型(LLM)。
Kore.ai	2024 年 1 月	英伟达	1.5 亿美元	Kore.ai 集成了英伟达的 GPU 加速卡 Riva，目前提供了一个企业无代码平台，帮助各种规模的公司以安全和负责任的方式与人工智能进行业务交互。

资料来源：半导体行业观察、Tech 商业，国盛证券研究所

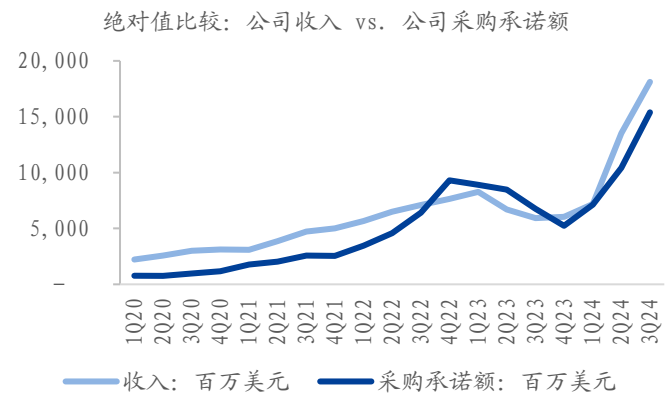
- **上游：加强对供应商的产能锁定**

在 CoWoS 和 HBM 产能稀缺的当前，“得产能者得天下”。英伟达除了此前与供应商建立的良好合作关系之外，亦在持续提升给供应商的采购协议金额。

截至 2024Q3 财季（2023 年 10 月），英伟达给供应商的采购承诺额已达 210 亿美金以上。根据我们的测算，其中后续 12 个月以内的采购承诺额在 150 亿美金以上。

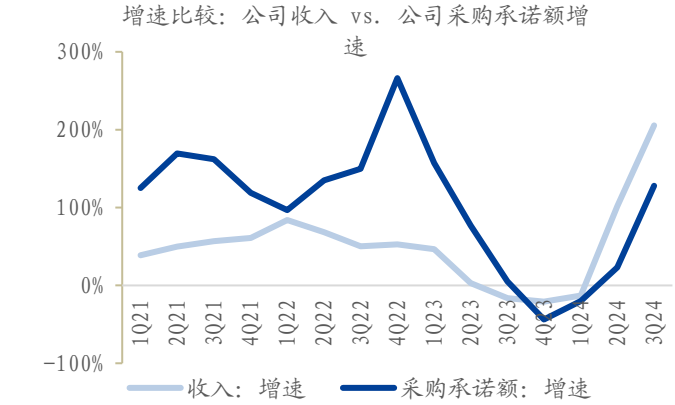
巨额采购协议的签订一定程度上帮助英伟达锁定了相应比例产能。而事实上，英伟达的收入兑现也和其上游采购承诺额呈现正相关——采购协议金额越高、说明其需求和产能保证度越高、收入也越高。

图表 34：绝对值：NVIDIA 收入 vs.采购承诺额（后续 12 个月）



资料来源：公司公告，国盛证券研究所

图表 35：增速：NVIDIA 收入 vs.采购承诺额（后续 12 个月）

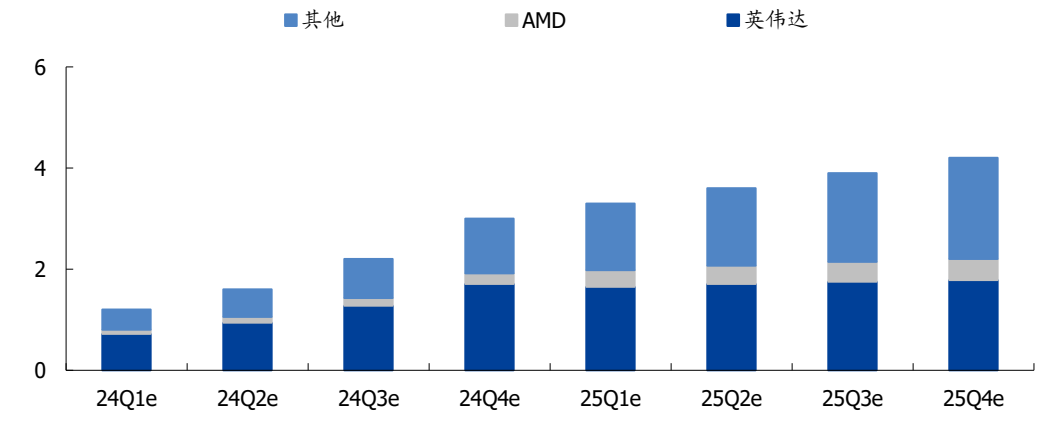


资料来源：公司公告，国盛证券研究所

基于 CoWoS 的产能增长、对英伟达不同产品线的产能分配等假设，按英伟达财年维度（例：FY 2025 财年为 2024 年 1 月至 2025 年 1 月），我们测算：

- H100 在 2025/2026 财年的出货量望达 209 万张/155 万张。
- H200 在 2025/2026 财年的出货量望达 35 万张/62 万张。
- B100 在 2025/2026 财年的出货量望达 23 万张/143 万张。

图表 36：CoWoS 产能及分配假设（万片 12 英寸晶圆）



资料来源：台湾电子时报、中关村在线，国盛证券研究所测算，注：此处为自然季度

图表 37：英伟达 H100、H200、B100 供给量测算

	24Q1e	24Q2e	24Q3e	24Q4e	25Q1e	25Q2e	25Q3e	25Q4e	26Q1e	26Q2e	26Q3e	26Q4e
英伟达 CoWoS（万片 12 英寸晶圆）	0.7	0.9	1.3	1.7	1.7	1.7	1.8	1.8	1.8	1.8	1.8	1.8
1)H100												
H100 CoWoS 占比	50%	50%	50%	50%	40%	35%	30%	25%	20%	20%	20%	20%
每片晶圆可切片数	30	30	30	30	30	30	30	30	30	30	30	30
H100 月供应量（万张）	11	14	20	27	20	18	16	13	11	11	11	11
H100 季供应量（万张）	32	42	59	81	59	54	47	40	32	32	32	32
2）H200												
H200 CoWoS 占比	0%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
每片晶圆可切片数	30	30	30	30	30	30	30	30	30	30	30	30
H200 月供应量（万张）	0	3	4	5	5	5	5	5	5	5	5	5
H200 季供应量（万张）	0	8	11	15	15	15	16	16	16	16	16	16
3）B100												
B100 CoWoS 占比	0%	0%	7%	27%	45%	47%	50%	55%	55%	55%	55%	55%
每片晶圆可切片数	14	14	14	14	14	14	14	14	14	14	14	14
B100 月供应量（万张）	0	0	1	6	10	11	12	14	14	14	14	14
B100 季供应量（万张）	0	0	3.8	19	31	34	37	41	41	41	41	41

资料来源：台湾电子时报、中关村在线，国盛证券研究所测算，注：此处为自然季度

4. 盈利预测、估值及投资建议

4.1 财务预测

1）数据中心业务

根据我们前述对英伟达数据中心 GPU 的需求和供给的假设和测算，我们估算:

- 2025 财年，公司 H100、H200、B100 出货量或达 209 万/35 万/23 万张。
- 2026 财年，公司 H100、H200、B100 出货量或达 155 万/62 万/143 万张。

基于此，我们估算英伟达的数据中心业务:

- 2025 财年，数据中心业务收入或达 819 亿美金，
- 2026 财年，数据中心业务收入或达 1087 亿美金。

2）游戏等其他业务

我们预计随着游戏 GPU 市场复苏以及英伟达产品的更新迭代，英伟达游戏业务在 2025 财年有望实现加速增长。专业可视化、智能驾驶等也有望保持稳健增长。

我们估算，公司整体业务的财务表现可达:

- 2025 财年，公司总收入或达 986 亿美金，GAAP 口径净利可达 483 亿美金，non-GAAP 口径净利可达 524 亿美金。
- 2026 财年，公司总收入或达 1282 亿美金，GAAP 口径净利可达 563 亿美金，non-GAAP 口径净利可达 610 亿美金。

综上，我们预计 2024-2026 财年公司 GAAP 净利润 284/483/563 亿美元，同比增长 549%/70%/16%。Non-GAAP 净利润 313/524/610 亿美元，同比增长 274%/68%/16%。

图表 38：英伟达财务预测：年度

百万美元	FY2020	FY2021	FY2022	FY2023	FY2024e	FY2025e	FY2026e
收入	10,918	16,675	26,914	26,974	59,368	98,607	128,150
数据中心	2,983	6,696	10,613	15,005	45,871	81,851	108,747
游戏	5,518	7,759	12,462	9,067	10,601	13,336	15,503
专业可视化	1,212	1,053	2,111	1,544	1,521	1,874	2,114
智能驾驶	700	536	566	903	1,081	1,184	1,344
其他	505	631	1,162	455	294	362	442
成本	-4,150	-6,279	-9,439	-11,618	-16,662	-28,495	-43,358
毛利	6,768	10,396	17,475	15,356	42,706	70,112	84,792
毛利率	62%	62%	65%	57%	72%	71%	66%
R&D rate	26%	24%	20%	27%	15%	14%	15%
S&G rate	10%	12%	8%	9%	5%	4%	4%
经营利润	2,846	4,532	10,041	4,224	31,101	52,139	60,697
经营利润率	26%	27%	37%	16%	52%	53%	47%
GAAP 净利	2,796	4,332	9,752	4,368	28,367	48,347	56,262
GAAP 净利润率	26%	26%	36%	16%	48%	49%	44%
Non-GAAP 净利	3,580	6,277	11,259	8,366	31,281	52,411	60,991
Non-GAAP 净利润率	33%	38%	42%	31%	53%	53%	48%

资料来源：公司公告，国盛证券研究所

图表 39：英伟达财务预测：季度

百万美元	FY1Q23	FY2Q23	FY3Q23	FY4Q23	FY1Q24	FY2Q24	FY3Q24	FY4Q24e
收入	8,288	6,704	5,931	6,051	7,192	13,507	18,120	20,549
数据中心	3,750	3,806	3,833	3,616	4,284	10,323	14,514	16,750
游戏	3,620	2,042	1,574	1,831	2,240	2,486	2,856	3,019
专业可视化	622	496	200	226	295	379	416	431
智能驾驶	138	220	251	294	296	253	261	271
其他	158	140	73	84	77	66	73	78
成本	-2,857	-3,789	-2,754	-2,218	-2,544	-4,045	-4,720	-5,353
毛利	5,431	2,915	3,177	3,833	4,648	9,462	13,400	15,196
毛利率	66%	43%	54%	63%	65%	70%	74%	74%
R&D rate	20%	27%	33%	32%	26%	15%	13%	13%
S&G rate	7%	9%	11%	10%	9%	5%	4%	4%
经营利润	1,868	499	601	1,256	2,140	6,800	10,417	11,744
经营利润率	23%	7%	10%	21%	30%	50%	57%	57%
GAAP 净利	1,618	656	680	1,414	2,043	6,188	9,243	10,893
GAAP 净利润率	20%	10%	11%	23%	28%	46%	51%	53%
Non-GAAP 净利	3,443	1,292	1,456	2,174	2,713	6,740	10,020	11,808
Non-GAAP 净利润率	42%	19%	25%	36%	38%	50%	55%	57%

资料来源：公司公告，国盛证券研究所

4.2 估值及投资建议

英伟达作为全球算力之源，我们看好其数据中心业绩的可持续性：

- **需求端：**英伟达数据中心业绩的可持续性，来自于 AI 算力需求的可持续性。1）训练端，更多国家和企业将入场 AI 军备战争，模型的参数数据量也更大。2）推理端，端侧 AI 的逐步落地、AI 应用向更多科技和制造领域破圈，均带来更强的推理算力需求。

  我们粗略测算：1）训练端：基于假设，至 2030 年全球累计需要相当于 2000 万张 H100 的等量算力需求。2）推理：基于假设，至 2030 年全球累计需要 1.16 亿张相当于 A30 的等量算力需求。英伟达作为全球算力之源，将充分受益。

- **供给端，**英伟达面对竞争亦做了充分的准备：1）软硬件产品上，公司在硬件产品上持续迭代新品，在软件架构 CUDA 上持续延续优势。2）上下游生态上，英伟达一方面通过投资参股等方式绑定下游企业的算力需求，一方面通过上百亿美金采购承诺额锁定上游产能。

- 基于 CoWoS 的产能增长、对英伟达不同产品线的产能分配等假设，我们测算：2025/2026 财年，H100 的出货量望达 209 万张/155 万张、H200 望达 35 万张/62 万张、B100 望达 23 万张/143 万张。

估值方面，纵向对比，英伟达当前估值处于历史均值以下。

横向对比，我们选取微软、Meta、AMD、超威电脑作为可比公司。其中微软和 Meta 既是英伟达 AI GPU 的重要客户，同时也在发力自研训练/推理芯片。AMD、超微电脑分别为 AI 算力芯片、AI 服务器重要标的。横向对比可比公司，我们认为英伟达作为高增长的全球算力龙头，有望享受估值溢价。

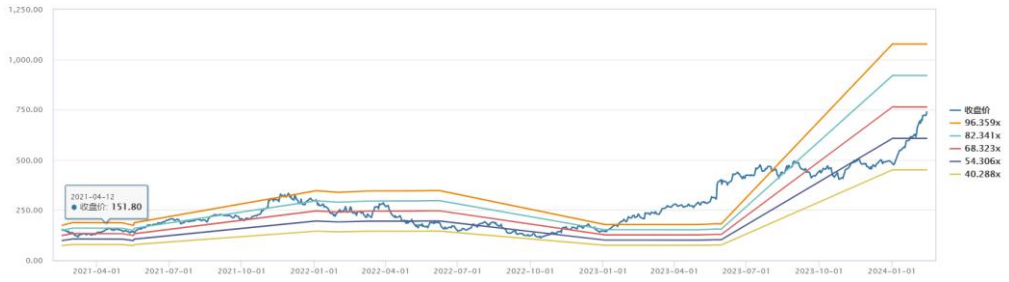
我们预计 2024-2026 财年公司 GAAP 净利润 284/483/563 亿美元，同比增长 549%/70%/16%。Non-GAAP 净利润 313/524/610 亿美元，同比增长 274%/68%/16%。考虑到英伟达净利润高速增长，我们认为英伟达合理市值为 20964 亿美元、对应股价为 840.6 美金，对应 40x FY2025e P/E（FY 2025 财年为 2024 年 1 月至 2025 年 1 月），首次覆盖给予“买入”评级。

图表 40：美股重点科技公司估值

股票简称	股票代码	收盘价（美元）	市值（亿美元）	EPS（美元）			P/E		
				FY2024E	FY2025E	FY2026E	FY2024E	FY2025E	FY2026E
微软	MSFT.O	406.3	30191.3	11.6	13.3	15.6	35.1	30.6	26.0
Meta	META.O	460.1	11730.3	19.1	21.9	26.1	24.1	21.0	17.6
AMD	AMD.O	171.5	2771.7	3.6	5.4	7.3	47.9	31.8	23.4
超微电脑	SMCI.O	791.5	442.7	21.6	28.2	32.9	36.6	28.0	24.1

资料来源：Bloomberg，国盛证券研究所。注：截止 2024 年 2 月 13 日收盘，EPS 来自 Bloomberg 一致预期

图表 41：英伟达 P/E band



资料来源：Wind，国盛证券研究所，注：截止 2024 年 2 月 13 日收盘，P/E 为 Wind 盈利预测

风险提示

下游 AI 应用不及预期。1）AI 硬件需求不及预期。AI 硬件相关行业涉及手机、PC、可穿戴设备、智能音箱、智能车、机器人等，若本身硬件行业如手机、PC 需求复苏不及预期，或 AI 功能对需求的提振不及预期，或将影响销售。2）AI 软件应用落地不及预期。如果大模型最终 B 端、C 端应用落地慢于预期，可能会带来相关公司基于 AI 的业务优化不及预期。

数据中心算力芯片竞争超预期。AI 芯片赛道中，除了种子选手英伟达之外，还有两类重要的阵营：1）以 AMD 和 Intel 为代表的 GPU 专业级新选手。2）以谷歌 TPU、微软 Athena 等为代表的云厂商自研芯片。若竞争超预期，或将影响英伟达的龙头地位。

AI 行业政策监管超预期。人工智能将对社会和经济产生深远影响，国内外均在审议出台 AI 相关监管法规制度。此外地缘政策也会影响 AI 算力的提供。若政策监管超预期，或将一定程度影响公司业务。

假设和测算误差风险。本文对全球企业在 AI 模型的训练和推理过程中所需要的算力芯片、CoWos 产能及分配、英伟达 GPU 供给量等测算均基于一系列假设，可能会与现实情况有所偏差，从而使得测算存在误差。

免责声明

国盛证券有限责任公司（以下简称"本公司"）具有中国证监会许可的证券投资咨询业务资格。本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告的信息均来源于本公司认为可信的公开资料，但本公司及其研究人员对该等信息的准确性及完整性不作任何保证。本报告中的资料、意见及预测仅反映本公司于发布本报告当日的判断，可能会随时调整。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司不保证本报告所含信息及资料保持在最新状态，对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司力求报告内容客观、公正，但本报告所载的资料、工具、意见、信息及推测只提供给客户作参考之用，不构成任何投资、法律、会计或税务的最终操作建议,本公司不就报告中的内容对最终操作建议做出任何担保。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

投资者应注意，在法律许可的情况下，本公司及其本公司的关联机构可能会持有本报告中涉及的公司所发行的证券并进行交易，也可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。

分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的任何观点均精准地反映了我们对标的证券和发行人的个人看法，结论不受任何第三方的授意或影响。我们所得报酬的任何部分无论是在过去、现在及将来均不会与本报告中的具体投资建议或观点有直接或间接联系。

投资评级说明

投资建议的评级标准		评级	说明
评级标准为报告发布日后的 6 个月内公司股价（或行业指数）相对同期基准指数的相对市场表现。其中 A 股市场以沪深 300 指数为基准；新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以摩根士丹利中国指数为基准，美股市场以标普 500 指数或纳斯达克综合指数为基准。	股票评级	买入	相对同期基准指数涨幅在 15%以上
		增持	相对同期基准指数涨幅在 5%~15%之间
		持有	相对同期基准指数涨幅在-5%~+5%之间
		减持	相对同期基准指数跌幅在 5%以上
	行业评级	增持	相对同期基准指数涨幅在 10%以上
		中性	相对同期基准指数涨幅在-10%~+10%之间
		减持	相对同期基准指数跌幅在 10%以上

国盛证券研究所

北京
地址：北京市东城区永定门西滨河路 8 号院 7 楼中海地产广场东塔 7 层
邮编：100077
邮箱：gsresearch@gszq.com

上海
地址：上海市浦明路 868 号保利 One56　1 号楼 10 层
邮编：200120
电话：021-38124100
邮箱：gsresearch@gszq.com

南昌
地址：南昌市红谷滩新区凤凰中大道 1115 号北京银行大厦
邮编：330038
传真：0791-86281485
邮箱：gsresearch@gszq.com

深圳
地址：深圳市福田区福华三路 100 号鼎和大厦 24 楼
邮编：518033
邮箱：gsresearch@gszq.com