

分析师：唐月
登记编码：S0730512030001
tangyue@ccnew.com 021-50586737

Sora 发布后，重新审视 AI 的发展走向

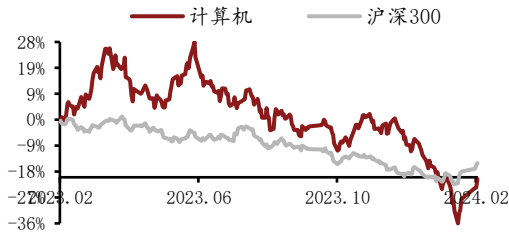
——人工智能专题研究（二）

证券研究报告-行业专题研究

强于大市(维持)

计算机相对沪深 300 指数表现

发布日期：2024 年 02 月 23 日



资料来源：中原证券，聚源

相关报告

- 《计算机行业年度策略：新格局下的新机遇》 2023-11-29
- 《计算机行业半年度策略：国产化、数字化、智能化》 2023-6-27
- 《人工智能专题研究（一）：ChatGPT 的出現有望加速我国在人工智能领域政策的出台》 2023-03-31

联系人：马嶽琦
电话： 021-50586973
地址： 上海浦东新区世纪大道 1788 号 16 楼
邮编： 200122

投资要点：

- **OpenAI 的重磅产品发布，再次吸引市场关注。**2月15日，openAI 发布了首个文生视频模型 Sora，其可以根据用户的文本提示快速制作长达一分钟的逼真视频。相比于 ChatGPT 的文字生成图片的功能来看，Sora 在对文字和图像识别和生成的基础上，又叠加了时间的维度，初步实现了对真实世界物理逻辑关系的理解。同时 Sora 的发布也意味着人类距离实现通用人工智能更进了一步。
- **大模型 AI 给我们带来的潜在变化包括：**
 - (1) 技术迭代速度明显加快。
 - (2) 算力和电力需求大幅提升。
 - (3) 带来降维打击，部分行业和个人竞争壁垒随之消亡。
 - (4) 或将带来第四次工业革命，推动国家之间竞争格局的重塑。
- 从芯片厂商最新财报情况来看，GPU 任然是数据中心的主角，但 AMD 让市场更加多元化。英伟达的数据中心业务在最新一个财季里再次超出市场预期，达到 **184.04 亿美元**，同比增速 **409%**，环比增长 **27%**。12 月 AMD 预测 AI 加速器市场规模在 2027 年将达到 4000 亿美元，较 2023 年 8 月 1500 亿美元的预期大幅提升。
- 继 2023 年 10 月美国商务部对 GPU 芯片禁令再一次升级以后，我们看到美国对于中国企业算力的限制还在加码过程中。在中国 GPU 供应面临障碍的同时，OpenAI 方面正在进行一个巨型的算力工程规划。面对当前的形式，国产 AI 芯片已经获得了更多厂商的青睐。考虑到当前全球政治、军事形势的复杂性，我们认为海外芯片及算力供应或将呈现收紧趋势。因而国产 GPU 芯片的生存环境将得到持续改善，后续一旦在生态和生产环节形成突破，将彻底改善其国内市场的竞争格局。
- 由于 AI 助力云迁移，Q4 前 4 大云计算厂商的收入增速都有加快迹象。由于微软、谷歌向 AI 发力，其资本开支都处于了历史高位水平，2023Q4 同比分别增长了 55%和 45%。从阿里、腾讯、百度三家国内互联网巨头来看，目前对于 AI 的资本投入较 2021 年历史高位仍有较大差距，较微软和谷歌在投入的力度和决心上仍然提升空间。

风险提示：国产芯片良率提升不及预期；人工智能技术推进不及预期。

内容目录

1. Sora 出现再次引起了人们对 AI 广泛的关注	3
2. 人工智能的深远影响	3
3. 英伟达销量爆增，AMD 大幅上调对 GPU 市场预期	5
4. 美国禁令加剧了国内 AI 芯片紧缺，国产芯片厂商获得更多青睐	7
5. AI 助力云迁移，谷歌、微软资本开支位居历史高位	8
6. 风险提示	10

图表目录

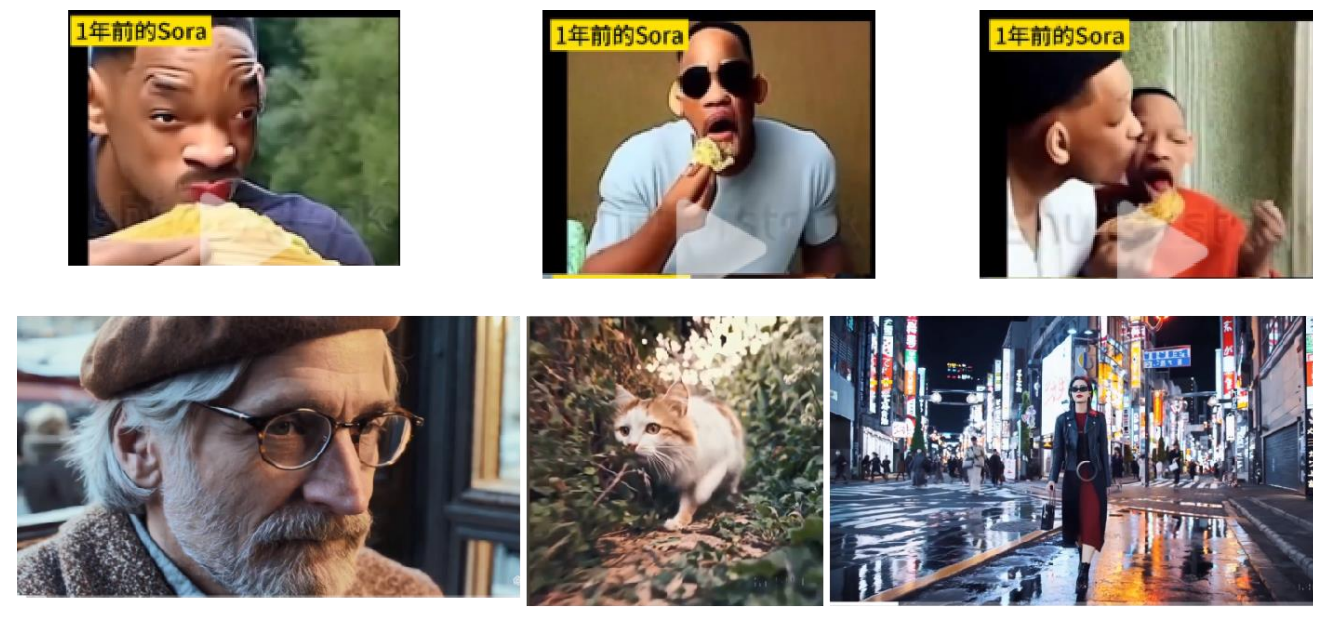
图 1: 当前 Sora 与 1 年前实现效果对比	3
图 2: 2022Q1-2023Q4 英伟达单季度收入的区域分布 (亿美元)	5
图 3: 人工智能产业架构	6
图 4: 2016Q1-2023Q4 三大芯片厂商数据中心业务收入及增速 (亿美元)	7
图 5: 2014Q1-2023Q4 全球前四大云计算厂商收入及增速 (亿美元)	8
图 6: 2019Q1-2023Q4 美股前五大科技企业资本开支 (百万美元)	9
图 7: 2015Q2-2023Q4 阿里、百度、腾讯单季度资本投入 (亿元)	9
表 1: OpenAI 的 GPT 大模型对比	3
表 2: 三次工业革命	4
表 3: 英伟达、AMD、华为人工智能芯片参数对比	5
表 4: 华为 AI 芯片应用的公开信息	8

1. Sora 出现再次引起了人们对 AI 广泛的关注

OpenAI 的重磅产品发布，再次吸引市场关注。2 月 15 日，openAI 发布了首个文生视频模型 Sora，其可以根据用户的文本提示快速制作长达一分钟的逼真视频。

对比 1 年前来看，Sora 当前视频的真实度和可用性取得了巨大的进步，到了可以以假乱真的程度。相比于 ChatGPT 的文字生成图片的功能来看，Sora 在对文字和图像识别和生成的基础上，又叠加了时间的维度，初步实现了对真实世界物理逻辑关系的理解。同时 Sora 的发布也意味着人类距离实现通用人工智能更进了一步。

图 1：当前 Sora 与 1 年前实现效果对比



资料来源：OpenAI，中原证券

2. 人工智能的深远影响

从深层次影响来看，大模型 AI 给我们带来的潜在变化可以概况成以下几个方面：

(1) 技术迭代速度明显加快。从 2023 年初，当 ChatGPT 火遍全网，市场还在质疑其对产业影响的落地，到 3 月 GPT-4 的惊艳亮相，到 11 月 GPT-4 Turbo 的性能提升和价格下降，到 2024 年 2 月 Sora 出现后惊叹技术迭代速度的加快和对技术落后的担忧，我们看到 OpenAI 一直在以超预期的方式刷新人们的认知。

表 1：OpenAI 的 GPT 大模型对比

模型	时间	参数	上下文限制	训练数据截止时间	公开的程度	备注
GPT-1	2018.6	1.17 亿			开源（技术细节）	
GPT-2	2019.2	15 亿			开源（源代码）	
GPT-3	2020.5	1750 亿		2019.10	论文	

GPT-3.5	2022.3	-	4K/16K	2021.6		2022.11 ChatGPT 基于 GPT 3.5 推出
GPT-4	2023.3	-	8K/32K	2021.9	测试结果	2023.7 向所有开发者开放使用权限
GPT-4 Turbo	2023.11	-	128K	2023.4		训练数据集 2023 年 4 月

资料来源：OpenAI，中原证券

(2) 算力和电力需求大幅提升。当前以大模型为支撑的人工智能，是建立在大力出奇迹的基础上的，也意味着对算力的需求呈现几何倍的增长。在这个过程中，GPU 取代 CPU 成为了数据中心的主流需求，同时带来数据中心设备的全面升级和对电力需求的全面提升。

(3) 带来降维打击，部分行业和个人竞争壁垒随之消亡。人工智能正在改变产业链的旧有格局，这意味着原有行业的竞争壁垒的消亡。因而我们也看到 Sora 发布以后，Adobe (ADBE.O) 股价出现了大幅的调整，体现出了市场对其发展前景的担忧。影视从业者 Sora 商用后，或将迎来职业危机。

由于大模型的模型智慧来源于对大量带标签的训练数据的认知，因而对 2C 领域的冲击更为显著。而 2B 领域由于数据的私密性，数据量相对有限，行业 knowhow 的壁垒更深。同时，人工智能在人类的已知领域将形成很好的替代，而对于未知领域的探索，人类还将占据主导。

对于计算机行业而言，大模型对通用工具的格局影响会大于对细分赛道的影响。随着大模型带来的代码能力的提升，人员效率也会得到极大的提升。

(4) 或将带来第四次工业革命，同时推动国家之间竞争格局的重塑。18 世纪英国通过蒸汽机的改良推动了机器的普及和大工厂制的建立，开创了机器替代手工劳动的时代，并成为了全球的主导力量；19 世纪六十年代以后，电机和内燃机的广泛应用，推动人类进入电气时代，让美国、德国崛起，英国丧失了垄断地位；20 世纪四五十年代以来，人类进入信息时代，美国在第三次工业革命中成为主导，处于霸主地位，发达国家和发展中国家的差距进一步拉大。

表 2: 三次工业革命

阶段	时间	标志	特点
第一次工业革命	1760-1840	蒸汽机	机器代替了手工劳动，人类从农耕文明向工业文明过渡
第二次工业革命	1860-1950	电机和内燃机	电力、钢铁、铁路、化工、汽车等重工业兴起，石油成为新能源
第三次工业革命	1950 至今	计算机	全球信息和资源交流变得更为迅速，大多数国家和地区都被卷入到全球化进程之中，世界政治经济格局进一步确立，人类文明的发达程度也达到空前的高度

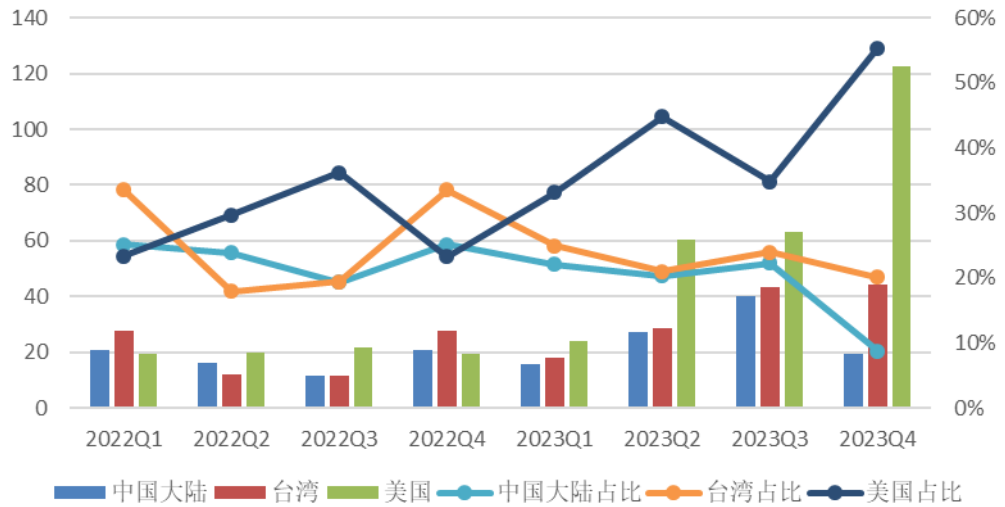
资料来源：百度百科，中原证券

在当前的人工智能浪潮中，美国以 OpenAI、谷歌为代表的企业的引领下，处于绝对的主导地位。我国虽然在人工智能领域也处于全球领先地位，但是较美国在大模型领域仍有较大差距。随着美国针对人工智能领域禁令的持续加码，我们在算力方面还面临着较大的发展瓶颈。

根据英伟达财报数据，随着最新禁令的出台，Q4 英伟达来自美国的收入占比进一步提升到了 55%，优势更加明显，而与此同时其来自中国大陆的收入从 Q3 的 40 亿美元降低到了

19 亿美元。

图 2：2022Q1-2023Q4 英伟达单季度收入的区域分布（亿美元）



资料来源：公司公告，Wind，中原证券（英伟达财报较正常季度滞后 1 个月）

3. 英伟达销量爆增，AMD 大幅上调对 GPU 市场预期

目前已经上市的 AI 芯片中，最强大的为 AMD 的 2023 年 12 月上市的 MI300X。而英伟达 2023 年 11 月发布的 H200 预计于 2024 年第二季度开始供货。H200 仍然采用 Hopper 架构，拥有 141GB 的内存，与前代产品 H100 相比，H200 的容量几乎翻了一番。根据界面新闻，由于市场对其 AI 芯片需求旺盛，英伟达将从每两年发布一次新架构转向每一年发布一次。英伟达还将于 2024 年发布基于 Blackwell 架构的 B100 芯片。

国产芯片方面，2023 年 10 月，科大讯飞表示昇腾 910B 能力已经基本做到可对标英伟达 A100。

表 3：英伟达、AMD、华为人工智能芯片参数对比

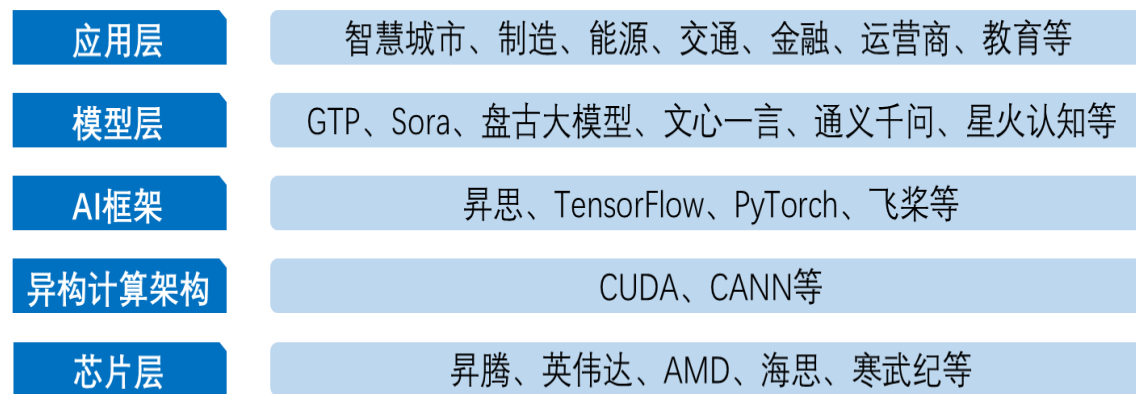
	A100 SXM	H100 SXM	H200 SXM	MI300X	Ascend 910
厂商	英伟达	英伟达	英伟达	AMD	华为
发布时间	2020.3	2022.3	2023.11	2023.6	2019.8
架构	Ampere	Hopper	Hopper	CDNA 3	达芬奇
工艺	7nm	4nm	4nm	5nm	7nm
晶体管数量	540亿	800 亿	800 亿	1530 亿	
最大功率	400W	700W		750W	310W
显存容量	80GB	80GB	141GB	192GB	32GB
显存带宽	2039GB/s	3.35TB/s	4.8TB/s	5.3TB/s	
显存颗粒	HBM2e	HBM3	HBM3e	HBM3	
传输速率	600GB/s	900GB/s			
FP64	9.7 TFLOPS	34 TFLOPS	34 TFLOPS		

FP64 Tensor Core	19.5 TFLOPS	67 TFLOPS	67 TFLOPS		
FP32	19.5 TFLOPS	67 TFLOPS	67 TFLOPS		
TF32	312TFIOPS	494.7 TFLOPS		653.7 TFLOPS	
TF32 Tensor Core		989 TFLOPS	989 TFLOPS		
BFLOAT16		989.4 TFLOPS		1307.4 TFLOPS	
BFLOAT16 Tensor Core	624 TFLOPS	1979 TFLOPS	1979 TFLOPS		
FP16		989.4 TFLOPS		1307.4 TFLOPS	256 TFLOPS
FP16 Tenor Core	624 TFLOPS	1979 TFLOPS	1979 TFLOPS		
FP8		1978.9 TFLOPS		2614.9 TFLOPS	
FP8 Tensor Core		3958 TFLOPS	3958 TFLOPS		
INT8		1978.92 TOPS		2614.92 TOPS	512 TOPS
INT8 Tensor Core	1248 TOPS	3958 TOPS	3958 TOPS		

资料来源：英伟达，AMD，华为，中原证券

除了硬件端的差距以外，华为和 AMD 这样后来者，在 CUDA 为代表的 GPU 相关软件生态方面与英伟达的差距更为明显，这也成为了当前制约其推广的核心难题。

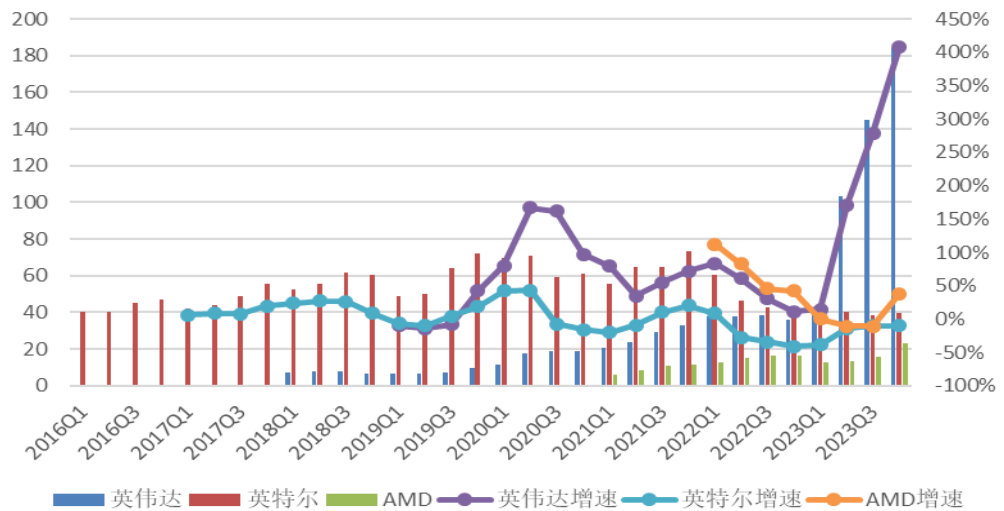
图 3：人工智能产业架构



资料来源：百度，华为，中原证券

从芯片厂商最新财报情况来看，GPU 任然是数据中心的主角，但 AMD 让市场更加多元化。英伟达的数据中心业务在最新一个财季里再次超出市场预期，达到 184.04 亿美元，同比增速 409%，环比增长 27%。得益于 AMD Instinct GPU 和 EPYC CPU 创纪录的季度销售额，其数据中心业务增长 38%。AI 领域发展速度超出 AMD 预期，其也将 2024 年数据中心 GPU 产品收入预期从 20 亿美元提升到了 35 亿美元。而与此同时，intel 数据中心业务面临更多的竞争压力，在 2023Q4 继续下滑 10%。

图 4：2016Q1-2023Q4 三大芯片厂商数据中心业务收入及增速（亿美元）



资料来源：上市公司财报，中原证券（英伟达财报较正常季度延后 1 个月）

2023 年 12 月 7 日，AMD 董事长苏姿丰在“Advancing AI”发布会上预测 AI 加速器市场规模在 2027 年将达到 4000 亿美元，较 2023 年 8 月 1500 亿美元的预期有了大幅的提升。

4. 美国禁令加剧了国内 AI 芯片紧缺，国产芯片厂商获得更多青睐

算力和芯片依然是制约 AI 发展的核心环节，继 2023 年 10 月美国商务部对 GPU 芯片禁令再一次升级以后，我们看到美国对于中国企业算力的限制还在加码过程中。1 月 27 日，美国商务部长 Gina Raimondo 宣布，美国政府正推出一项提案，阻止外国实体，特别是来自中国的实体，使用美国的云计算进行人工智能模型的训练。根据 2 月 1 日快科技消息，刚退市的“自动驾驶卡车第一股”图森未来，计划向澳大利亚运送 24 台英伟达 A100 GPU 芯片，却被美国拦截，美国政府担忧这些高性能 GPU 芯片可能会被转售到中国。

在 2023 年 10 月禁令后，英伟达又对华推出 3 款“特供版”AI 芯片，包括了分别针对训练、推理、边缘场景的 HGX H20、L20 PCIe 和 L2 PCIe，其中最强的 H20 整体算力要比英伟达 H100 GPU 芯片降 80%左右。根据财联社 2 月 4 日消息，英伟达 H20 的终端产品已可接受预订。互联网大厂测试结果不理想，订购数量将远少于原计划。

在中国 GPU 供应面临障碍的同时，OpenAI 方面正在进行一个巨型的算力工程规划。根据华尔街见闻 2 月 9 日消息，Sam Altman 计划筹集数 5-7 万亿美元，重塑半导体产业格局，包括增加芯片、能源和数据中心的全球基础设施和供应链等方面。

面对当前的形式，国产 AI 芯片已经获得了更多厂商的青睐。根据 IT 时报的报道，基于华为昇腾 910B 芯片的全套国产智算集群，正在国内多个智算中心、大模型厂商、电信运营商处进行测试，尽管在通用性上与 A100 仍有差距，但调优之后，在某些模型上与 A100 几乎可以

比肩。根据路透社 2 月 5 日消息，由于市场对华为昇腾芯片需求的飙升和产能的限制，华为放缓了用于 Mate 60 手机的麒麟芯片的生产，转产昇腾芯片，但是目前产量还受到良率的限制。

表 4: 华为 AI 芯片应用的公开信息

新闻时间	用户	概况
2024.2.18	科大讯飞	星火大模型 V3.5 基于科大讯飞联合华为打造的全国产化算力底座“飞星一号”平台，该平台采用的是华为的昇腾 910B。
2023.11.9	360	360 集团创始人周鸿祎在 2023 年世界互联网大会乌镇峰会上提到，360 也采购了华为 1000 片左右的 AI 芯片。
2023.11.7	百度	8 月百度采购了华为 1600 颗昇腾 910B AI 芯片，到 10 月份已经交付了其中的 60%，大概约 1000 颗。
2023.10.13	中国电信	采购鲲鹏芯片的训练型 AI 算力服务器 1977 台。

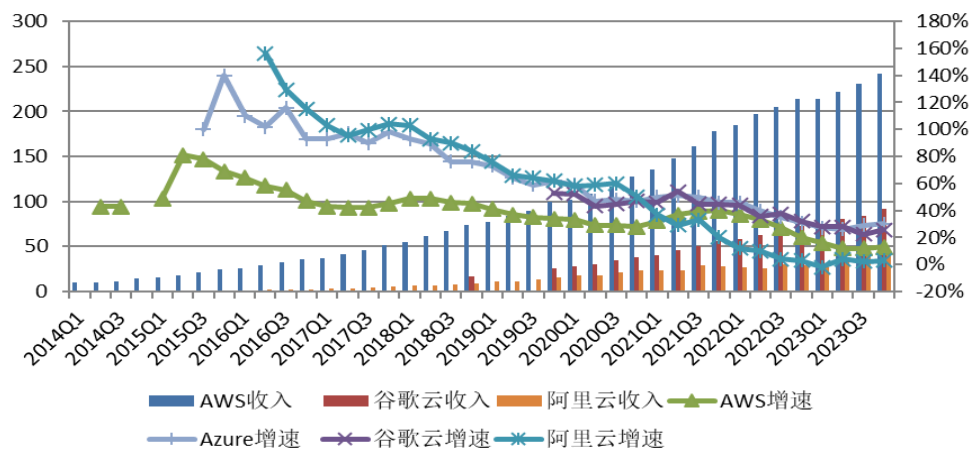
资料来源：金融界，快科技，智通财经，中原证券

考虑到当前全球政治、军事形势的复杂性，我们认为海外芯片及算力供应或将继续呈现收紧趋势。因而国产 GPU 芯片的生存环境将得到持续改善，后续一旦在生态和生产环节形成突破，将彻底改善其国内市场的竞争格局。

5. AI 助力云迁移，谷歌、微软资本开支位居历史高位

AI 助力云迁移，Q4 全球前 4 大云计算厂商的收入增速都有加快迹象。其中 AWS 收入增速从上各个季度的 12% 增加到 13%；Azure 增速继续 Q3 增加 3pct 以后，Q4 又继续提升 1pct 至 30%，继续实现引领；谷歌云增速从 Q3 的 22.5% 提升到 25.7%，增长显著；阿里较 Q3 提升 1pct 至 3%。

图 5: 2014Q1-2023Q4 全球前四大云计算厂商收入及增速（亿美元）

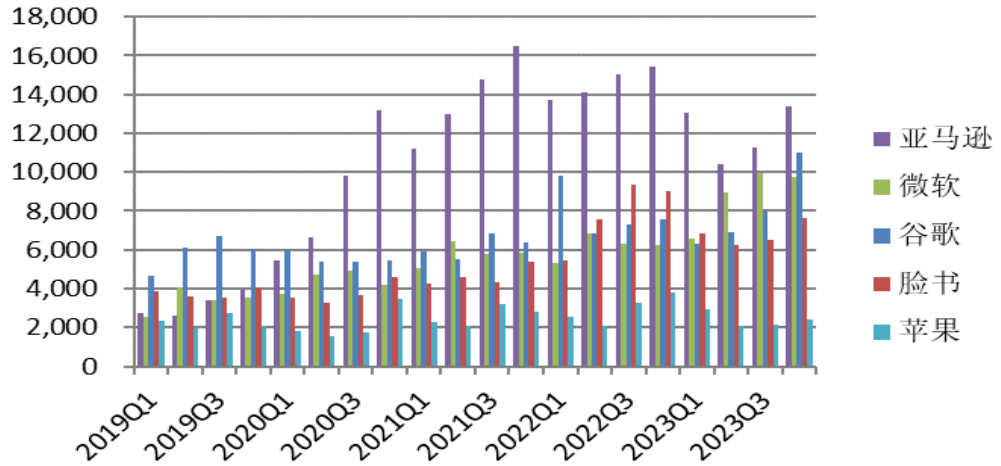


资料来源：上市公司财报，中原证券

5 大科技厂商总体显示出资本开支加快迹象，微软、谷歌向 AI 发力。其中微软的资本开支在 2023 整体呈现逐季增长的趋势，2023Q4 较 Q3 略有回落，但仍然维持在历史的高位水

平，同比增长 55%；谷歌呈现出追赶之势，2023Q1 以后资本开支快速增长，在 Q4 重新超过微软到达历史高位，同比增长 45%；亚马逊、Meta、Apple 本轮资本开支低点出现在 2022Q2，而后逐季回升，但是同比仍然分别下滑了 13%、15%和 37%。

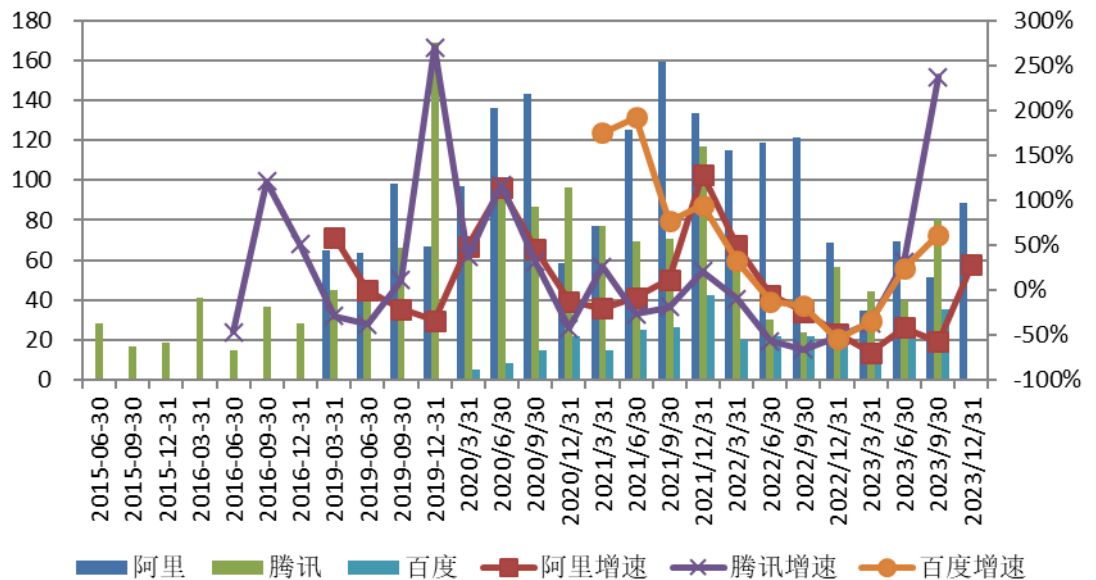
图 6：2019Q1-2023Q4 美股前五大科技企业资本开支（百万美元）



资料来源：Wind，中原证券

从国内互联网巨头来看，目前对于 AI 的资本投入力度和决心上较微软和谷歌而言仍然有待提升。2023Q4，阿里资本投入 89 亿元，同比增长 28.4%，是 2021Q3 峰值的 56%；腾讯 2023Q3 资本投入 80 亿元，同比增长 236.8%，是 2021Q4 峰值的 69%；百度 2023Q3 资本开支 35.29 亿元，同比增长 61.3%，是 2021Q4 峰值的 83%，是三家投入积极性最高的。

图 7：2015Q2-2023Q4 阿里、百度、腾讯单季度资本投入（亿元）



资料来源：上市公司财报，中原证券

在我国 AI 技术落后的情况下，我们也期待国内科技巨头在 AI 领域加大投入力度，探寻赶超的机会。

6. 风险提示

国产芯片良率提升不及预期；人工智能技术推进不及预期。

行业投资评级

强于大市：未来 6 个月内行业指数相对沪深 300 涨幅 10% 以上；

同步大市：未来 6 个月内行业指数相对沪深 300 涨幅-10% 至 10% 之间；

弱于大市：未来 6 个月内行业指数相对沪深 300 跌幅 10% 以上。

公司投资评级

买入：未来 6 个月内公司相对沪深 300 涨幅 15% 以上；

增持：未来 6 个月内公司相对沪深 300 涨幅 5% 至 15%；

谨慎增持：未来 6 个月内公司相对沪深 300 涨幅-10% 至 5%；

减持：未来 6 个月内公司相对沪深 300 涨幅-15% 至 -10%；

卖出：未来 6 个月内公司相对沪深 300 跌幅 15% 以上。

证券分析师承诺

本报告署名分析师具有中国证券业协会授予的证券分析师执业资格，本人任职符合监管机构相关合规要求。本人基于认真审慎的职业态度、专业严谨的研究方法与分析逻辑，独立、客观的制作本报告。本报告准确的反映了本人的研究观点，本人对报告内容和观点负责，保证报告信息来源合法合规。

重要声明

中原证券股份有限公司具备证券投资咨询业务资格。本报告由中原证券股份有限公司（以下简称“本公司”）制作并仅向本公司客户发布，本公司不会因任何机构或个人接收到本报告而视其为本公司的当然客户。

本报告中的信息均来源于已公开的资料，本公司对这些信息的准确性及完整性不作任何保证，也不保证所含的信息不会发生任何变更。本报告中的推测、预测、评估、建议均为报告发布日的判断，本报告中的证券或投资标的价格、价值及投资带来的收益可能会波动，过往的业绩表现也不应当作为未来证券或投资标的表现的依据和担保。报告中的信息或所表达的意见并不构成所述证券买卖的出价或征价。本报告所含观点和建议并未考虑投资者的具体投资目标、财务状况以及特殊需求，任何时候不应视为对特定投资者关于特定证券或投资标的的推荐。

本报告具有专业性，仅供专业投资者和合格投资者参考。根据《证券期货投资者适当性管理办法》相关规定，本报告作为资讯类服务属于低风险（R1）等级，普通投资者应在投资顾问指导下谨慎使用。

本报告版权归本公司所有，未经本公司书面授权，任何机构、个人不得刊载、转发本报告或本报告任何部分，不得以任何侵犯本公司版权的其他方式使用。未经授权的刊载、转发，本公司不承担任何刊载、转发责任。获得本公司书面授权的刊载、转发、引用，须在本公司允许的范围内使用，并注明报告出处、发布人、发布日期，提示使用本报告的风险。

若本公司客户（以下简称“该客户”）向第三方发送本报告，则由该客户独自为其发送行为负责，提醒通过该种途径获得本报告的投资者注意，本公司不对通过该种途径获得本报告所引起的任何损失承担任何责任。

特别声明

在合法合规的前提下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问等各种服务。本公司资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告意见或者建议不一致的投资决策。投资者应当考虑到潜在的利益冲突，勿将本报告作为投资或者其他决定的唯一信赖依据。