



AI 周观察

行业深度研究(深度)
证券研究报告

分析师: 刘道明 (执业 S1130520020004) 联系人: 黄晓军 (执业 S1130122050092) 联系人: 麦世学 (执业 S1130123100111)
liudaoming@gjzq.com.cn huangxiaojun@gjzq.com.cn maishixue@gjzq.com.cn

本周英伟达财报发布, 关注毛利率和后续产品迭代情况

报告摘要

- OpenAI 下一代人工智能模型 (Orion) 在性能提升方面出现放缓, 未达到业界预期, 训练初期就已与 GPT-4 相当, 但随着训练深入, 性能提升减弱。这可能意味着扩展规律 (通过增加模型规模和数据量提升性能) 不再适用, 可能影响 GPGPU 市场需求。企业将重新评估计算资源投资, 可能更加注重算法优化和模型架构创新, 推动对小规模、高效率计算集群的需求增加。
- AI 聊天助手活跃度仍在上升, OpenAI 的 Operator 实现类似 “Computer Use” 功能, 将在明年初推出, AI 代理效果将持续提升至可用程度, 对传统交互逻辑有较大的影响。
- SK 海力士在 SK AI Summit 2024 上宣布正在开发容量为 48GB 的 HBM3e 16hi 产品, 预计于 2025 年上半年送样。该产品将提供更高的位元容量, 并在 HBM4 世代量产前提前提升性能。与 HBM3e 8hi 和 12hi 产品相比, 16hi 产品将采用先进的 MR-MUF 堆栈制程, 能够提供更高的堆栈层数和计算频宽。此外, SK hynix 通过提前量产 HBM3e 16hi, 将积累经验, 为未来 HBM4 16hi 的量产奠定基础, 并可能推出混合键合制程版本, 扩展高运算频宽应用。
- 近期, DRAM 和 NAND 存储现货市场价格持续下跌, 近八周环比下滑, 10 月业绩显示多家厂商单月营收低于去年同期。市场情绪悲观, 部分厂商预期 NAND 原厂可能在 12 月减产, 推高 2025 年下半年供应紧缺的风险。然而, 市场尚处于观望阶段, 部分原厂未明确行动。在 PC 终端市场, 尽管推出了 AI PC 产品, 但 AI PC 的实际出货量渗透率仍较低, 未能显著提振整体业绩。预计第四季度 PC 市场需求仍不明朗, 可能面临 “旺季不旺” 局面。
- 2024 年三季度, 全球 AI 笔电销量约为 1450 万台, 其中英特尔核 AI 笔电销量约为 720 万台, 占比约 50%, 苹果、AMD、高通核的 AI 笔电销量分别为 540 万、134 万、55 万台, 占比分别为 37%、9%、4%。我们认为未来 AI 笔电 SoC/CPU 的竞争将会更为激烈。
- 英伟达将于下周三发布 2025 财年第三季度财报, 重点关注 Blackwell 产品的出货情况和对毛利率的影响。尽管云服务厂商的资本开支增速放缓, 但主要厂商仍预计持续增长, Blackwell 产品需求不应成问题。市场预计第三季度毛利率为 75.01%, 且未来几季度不会显著回升, 反映出产品迭代难度和研发投入增加。预计本季度营业收入为 329.14 亿美元, 同比增长约 80%。然而, 产能扩张的局限性和前期高基数影响可能导致收入增速放缓。此外, 市场尚未充分考虑到 Rubin 下一代产品可能的延迟或性能提升不及预期, 需关注管理层对未来产品的相关描述。

风险提示

- 芯片制程发展与良率不及预期
- 中美科技领域政策恶化
- 智能手机销量不及预期



内容目录

财报日历与前瞻.....	3
AI 模型与应用动态.....	4
AI 聊天助手应用热度持续增长.....	5
视频生成应用发展需要 Sora 和其他新模型的发布助力.....	6
OpenAI 下一代模型性能提升放缓，或将影响 GPGPU 采购决策.....	6
存储市场动态.....	6
SK 海力士追加推出 HBM3e 16hi 产品.....	6
存储现货市场持续走跌，AIPC 拉货效果不及预期.....	7
AI PC 销量.....	8
英特尔在 24Q3AI 笔电中表现良好.....	8
财报前瞻.....	9
英伟达：关注 Blackwell 首个交付季度表现.....	9
风险提示.....	10



财报日历与前瞻

图表1: 美股 AI 相关季报日历与一致预期

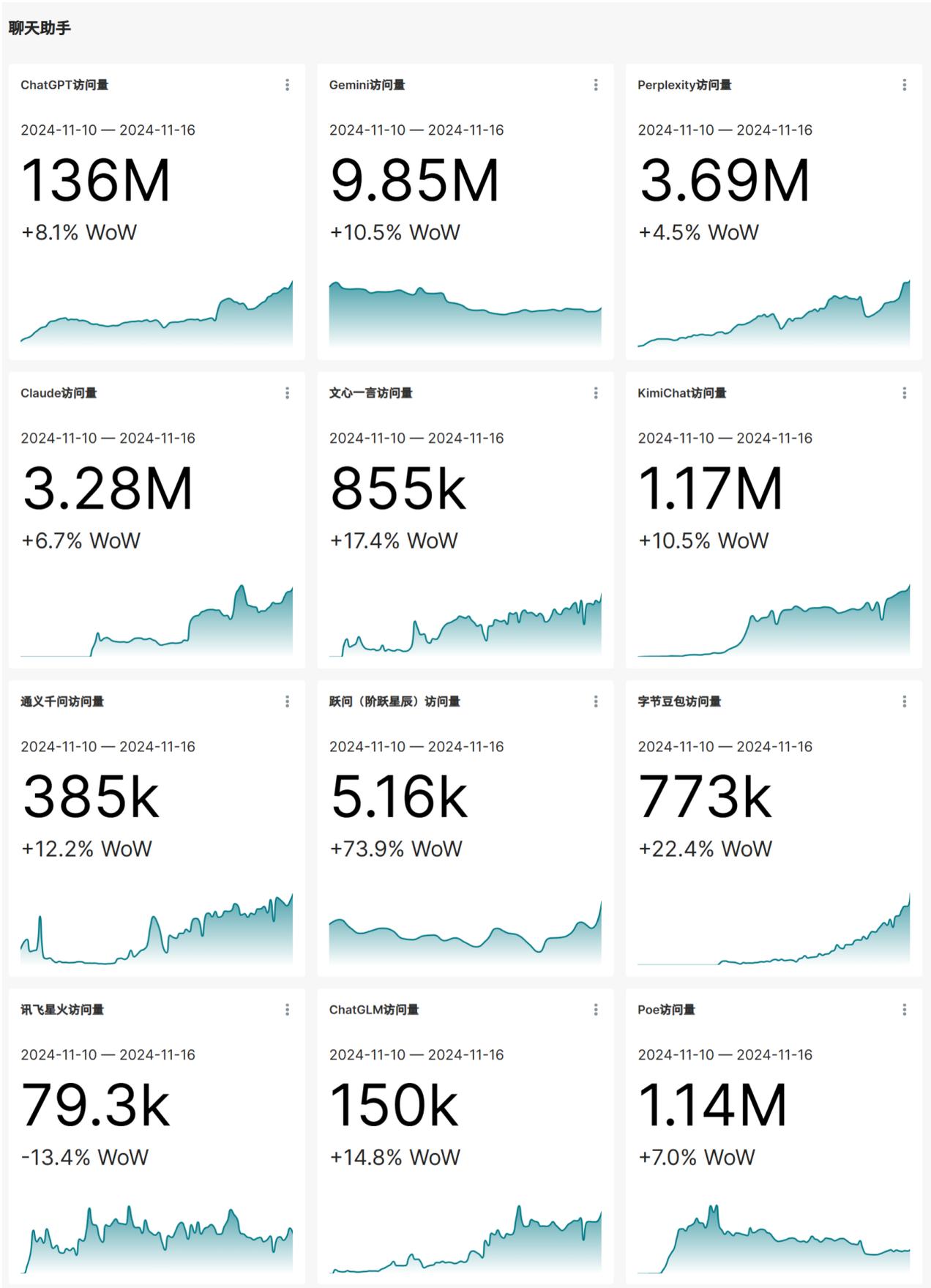


来源: Reuters、国金证券研究所



AI 模型与应用动态

图表2: 聊天助手类 AI 应用日活跃度





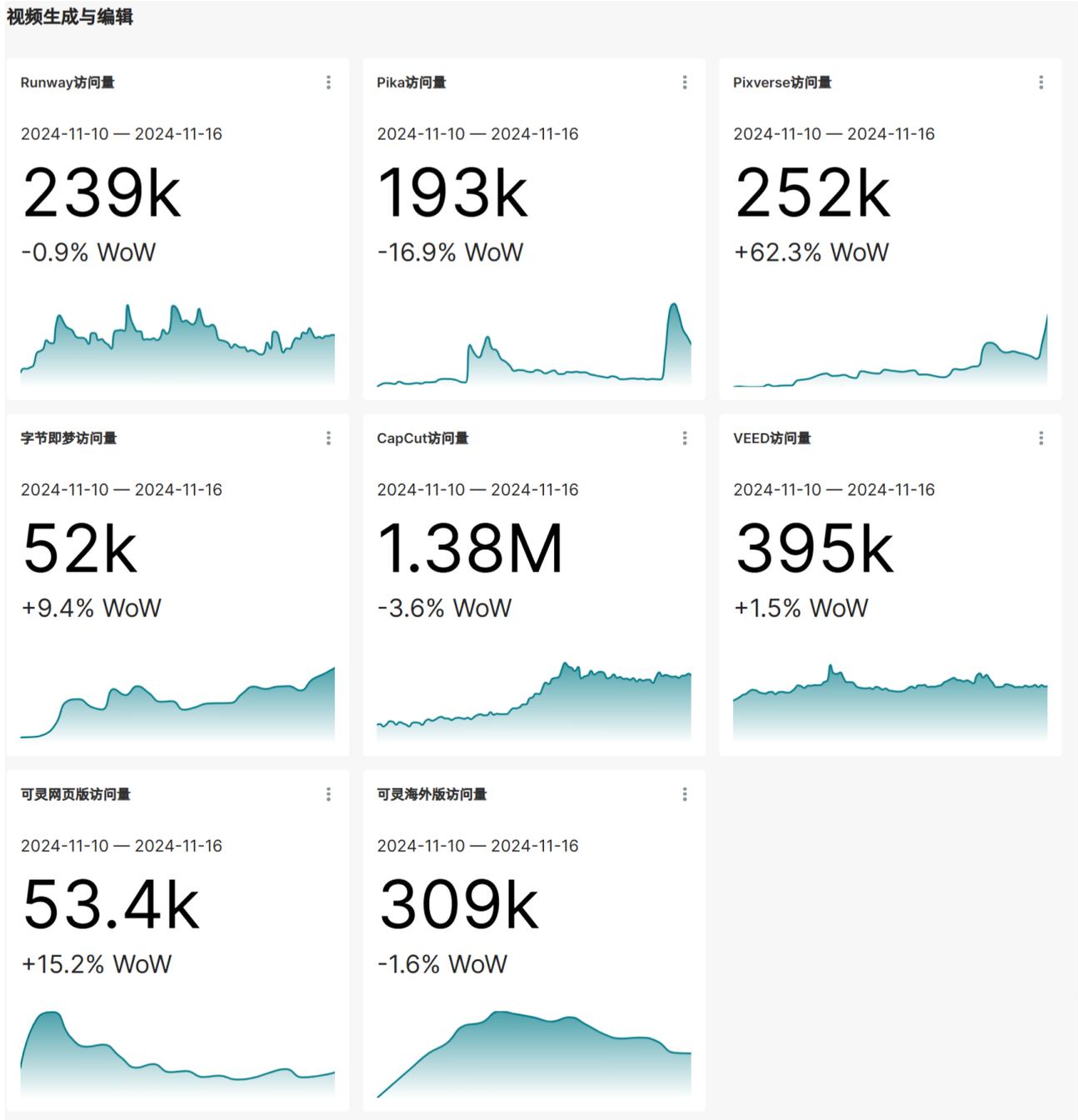
来源: SimilarWeb、国金证券研究所

AI 聊天助手应用热度持续增长

从 AI 应用的日活跃度数据看, ChatGPT 活跃度再创新高, 周均环比增长 8.1%, 国内大多数应用的活跃度也在快速提升。

在基础模型领域, Bloomberg 透露 OpenAI 计划于明年 1 月推出代号为 "Operator" 的 AI 代理工具, 该工具将通过 API 形式向开发者开放。Operator 具备代码编写和差旅预订等实际操作能力, 标志着 AI 在实际应用场景中的进一步深化。上周, 阿里云发布 Qwen2.5-Coder 系列模型, 该系列覆盖 0.5B 至 32B 等六种规格, 其中 32B 版本已达到开源代码模型的 SOTA 水平, 展现出与 GPT-4o 相媲美的编程能力。

图表3: 视频类 AI 应用日活跃度



来源: SimilarWeb、国金证券研究所



视频生成应用发展需要 Sora 和其他新模型的发布助力

视频生成应用的日活跃度的增速明显下滑，多数与上周环比微增或者下滑。Pixverse 新发模型后活跃度仍在快速增长，上周环比上升 62%。

在视频生成领域，Vidu 发布 1.5 版本大模型，在多维度实现技术突破。新版本在复杂主体控制、人物面部特征连贯性以及多主体交互等方面均取得显著进展。该版本展现出与大语言模型类似的上下文理解和记忆能力，反映出视觉模型在系统训练后可实现深度认知，这一现象或将为视频生成领域带来范式转变。

OpenAI 下一代模型性能提升放缓，或将影响 GPGPU 采购决策

近期有消息指出，OpenAI 的下一代人工智能模型在性能提升方面出现了放缓的趋势，未能达到业界的预期。据内部测试显示，代号为 Orion 的新模型在仅完成 20% 的训练时，就已经达到了与 GPT-4 相当的性能水平。然而，令人意外的是，随着训练的深入，模型性能的提升幅度却逐渐减小。相比于从 GPT-3 到 GPT-4 所取得的显著进步，GPT-4 到 GPT-5 的性能提升似乎并不明显。

这一现象引发了业界对 Scaling Law (扩展规律) 可能失效的担忧。一直以来，扩展规律是推动大型语言模型性能提升的关键理论，认为通过增加模型的参数规模和训练数据量，模型的性能将持续增强。然而，如果这一规律在当前模型规模下不再适用，意味着单纯依靠扩大参数和数据量可能无法继续带来预期的性能提升。

这种趋势可能对市场产生深远的影响，特别是在 GPGPU(通用图形处理器)采购决策方面。企业一直依赖于提升计算能力来支持更大规模的模型训练，如果模型性能不再随规模扩展而显著提升，那么对超大规模集群的需求可能不会如之前预期的那样强劲。企业可能需要重新评估其计算资源的投资策略，转而关注算法优化、模型架构创新或其他提高模型性能的方法。

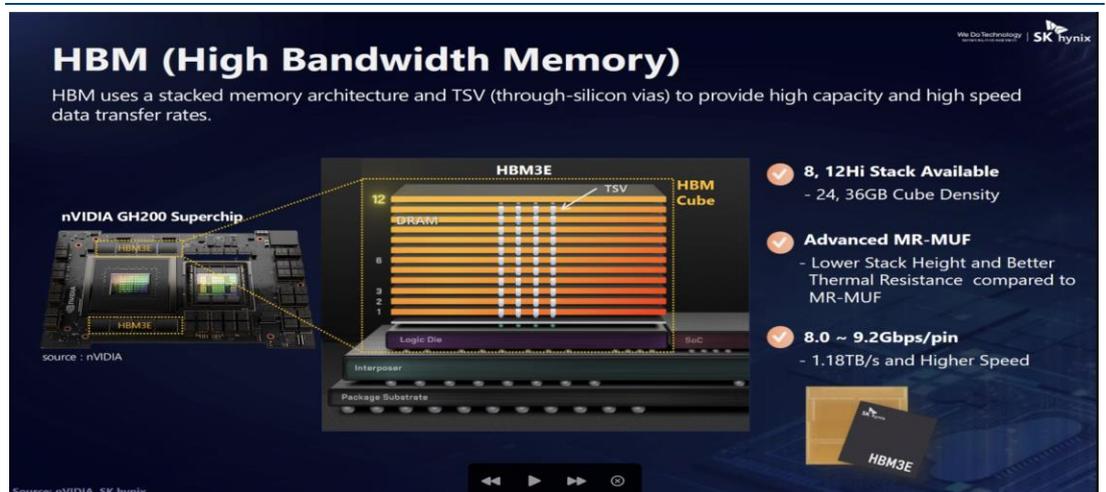
总的来说，OpenAI 新一代模型性能提升的放缓，可能预示着人工智能领域需要探索新的发展路径。业界需要重新审视现有的理论和实践，寻找突破当前瓶颈的创新方法，以继续推动人工智能技术的进步。同时，这也可能促使企业在计算资源的投入上更加谨慎，推动市场对小规模、高效率计算集群的需求增加。

存储市场动态

SK 海力士追加推出 HBM3e 16hi 产品

SK hynix (SK 海力士) 近日在 SK AI Summit 2024 活动中透露，正在开发每颗容量为 48GB 的 HBM3e 16hi 产品，预计于 2025 年上半年送样。这款新产品的潜在应用包括云端服务业者(CSP)自行研发的 ASIC 和通用型 GPU，有望在 HBM4 世代量产前，于 HBM3e 世代提前提升位元容量上限。通常，HBM 供应商在各世代会推出两种不同堆栈层数的产品，如 HBM3e 的 8hi 和 12hi，HBM4 的 12hi 和 16hi。

图表4: HBM3e 先前仅有 8hi 和 12hi 两种堆栈层数



来源: eetimes、国金证券研究所



在各业者计划于 2025 年下半年投入 HBM4 12hi 的情况下, SK hynix 选择在 HBM3e 追加推出 16hi 产品, 主要原因包括: 首先, 台积电(TSMC)的 CoWoS-L 封装尺寸将于 2026 至 2027 年间扩大, 每个系统级封装(SiP)可搭载的 HBM 数量增加, 但后续升级仍面临技术挑战和不确定性, 因此在量产难度更高的 HBM4 16hi 之前, 先提供 HBM3e 16hi 作为大位元容量的选择; 其次, 在 HBM3e 12hi 升级至 HBM4 12hi 过程中, HBM3e 16hi 可作为提供低 IO 数、小晶粒尺寸和大位元容量的选项。SK hynix 的 HBM3e 16hi 将采用 Advanced MR-MUF 堆栈制程, 与 TC-NCF 制程相比, 更易实现高堆栈层数和高运算频宽。由于 HBM4 及 HBM4e 世代皆预计设计 16hi 产品, SK hynix 率先量产 HBM3e 16hi, 不仅可及早累积量产经验, 加速后续 HBM4 16hi 的量产进程, 也不排除未来推出混合键合(hybrid bonding)制程版本的 16hi 产品, 以拓展更高运算频宽的应用客户群。

存储现货市场持续走跌, AIPC 拉货效果不及预期

从周度价格涨跌幅来看, DRAM 和 NAND 现货价近八周环比持续走跌, 10 月业绩显示, 多家厂商的单月营收低于去年同期水平, 已连续两个月同比下滑, 市场悲观情绪蔓延。群联电子 CEO 潘健成预估, NAND 原厂很可能将在今年 12 月减产, 推动 2025 年下半年再次出现供应紧缺的局面。然而, 目前部分原厂尚未明确行动, 供需双方均在观望市场变化。PC 终端方面, 近期华硕、宏碁等 PC 品牌厂商的最新业绩出现下滑, 10 月营收环比下降近 20% 至 30%, 创下近一个季度新低。尽管今年以来 PC OEM 积极推出 AI PC 产品, 但实际 AI PC 出货量渗透率仍然较低, 对 PC 品牌厂商整体业绩的贡献有限。对于后市, 部分 PC OEM 厂商预计第四季度 PC 市场整体需求不明朗, 可能面临“旺季不旺”的局面。

图表5: DRAM Wafer 现货价近八周涨跌幅

	2024-39	2024-40	2024-41	2024-42	2024-43	2024-44	2024-45	2024-46
DRAM:DDR3 2Gb 128Mx16 1600/1866	-2.41%	0.00%	-1.23%	-1.25%	0.00%	0.00%	-1.27%	0.00%
DRAM:DDR3 2Gb 256Mx8 1600/1866	-1.96%	-2.00%	0.00%	-1.02%	-3.09%	-2.13%	-2.17%	-1.11%
DRAM:DDR3 4Gb 256Mx16 1600/1866	-2.27%	-1.16%	0.00%	-2.35%	-1.20%	-2.44%	-2.50%	-1.28%
DRAM:DDR3 4Gb 512Mx8 eTT	-5.00%	0.00%	-5.26%	0.00%	5.56%	0.00%	5.26%	0.00%
DRAM:DDR4 16Gb(1Gx16)3200	-1.52%	-0.62%	-0.62%	0.00%	0.00%	0.31%	-1.24%	-1.89%
DRAM:DDR4 16Gb(2Gx8)2666 Mbps	-0.26%	0.00%	0.00%	0.00%	0.00%	0.26%	-0.26%	-0.26%
DRAM:DDR4 16Gb(2Gx8)3200	-0.61%	-0.31%	0.00%	-0.31%	0.00%	0.00%	0.00%	-0.93%
DRAM:DDR4 16Gb(2Gx8)eTT Mbps	-3.03%	-2.73%	-2.41%	-3.29%	-2.55%	-2.62%	-2.24%	-2.75%
DRAM:DDR4 4Gb 512Mx8 eTT	-1.67%	-1.69%	-1.72%	0.00%	-1.75%	0.00%	0.00%	0.00%
DRAM:DDR4 4Gb(256Mx16)2400/2666	-1.54%	0.00%	0.00%	-0.83%	-1.68%	-0.85%	-0.86%	0.00%
DRAM:DDR4 4Gb(512Mx8)2400/2666	-0.81%	0.00%	0.00%	0.00%	0.00%	-0.81%	0.00%	0.00%
DRAM:DDR4 8Gb(1Gx8)2666 Mbps	-1.02%	-0.62%	0.00%	-1.04%	-0.52%	-1.58%	-1.60%	0.00%
DRAM:DDR4 8Gb(1Gx8)3200	-1.16%	-1.18%	-1.19%	-1.81%	-0.61%	-0.62%	-1.86%	-1.27%
DRAM:DDR4 8Gb(1Gx8)eTT	-2.02%	-2.06%	-2.11%	-2.15%	-3.30%	-1.14%	-2.30%	-2.35%
DRAM:DDR4 8Gb(512Mx16)2666 Mbps	-1.16%	-1.75%	-0.60%	-1.20%	-0.61%	-1.22%	-0.62%	-0.62%
DRAM:DDR4 8Gb(512Mx16)3200	-0.60%	-0.60%	0.00%	-0.60%	0.00%	0.00%	-0.61%	0.00%
DRAM:DDR5 16G(2Gx8)4800/5600	1.01%	-0.60%	-0.81%	-0.61%	-2.05%	0.00%	0.42%	0.00%

来源: dramexchange、国金证券研究所

图表6: NAND Wafer 现货价近八周涨跌幅

	2024-39	2024-40	2024-41	2024-42	2024-43	2024-44	2024-45	2024-46
MLC 128Gb	-0.15%	-0.15%						
MLC 256Gb	-0.27%	-0.18%						
MLC 32Gb	1.42%	2.34%						
MLC 64Gb	0.73%	0.24%						
QLC 1Tb	0.00%		-8.62%	0.00%	-1.89%	-1.92%	-1.96%	0.00%
SLC 16Gb	1.96%	2.56%						
SLC 1Gb	1.27%	0.00%						
SLC 2Gb	0.84%	0.00%						
SLC 4Gb	0.00%	0.00%						
SLC 8Gb	0.00%	1.39%						
TLC 1Tb	0.00%		-7.35%	0.00%	-1.59%	-3.23%	-1.67%	0.00%
TLC 256Gb	0.00%		0.00%	-12.50%	0.00%	0.00%	0.00%	0.00%
TLC 512Gb	0.00%		-8.11%	0.00%	-2.94%	-3.03%	0.00%	0.00%

来源: dramexchange、中国闪存市场、国金证券研究所

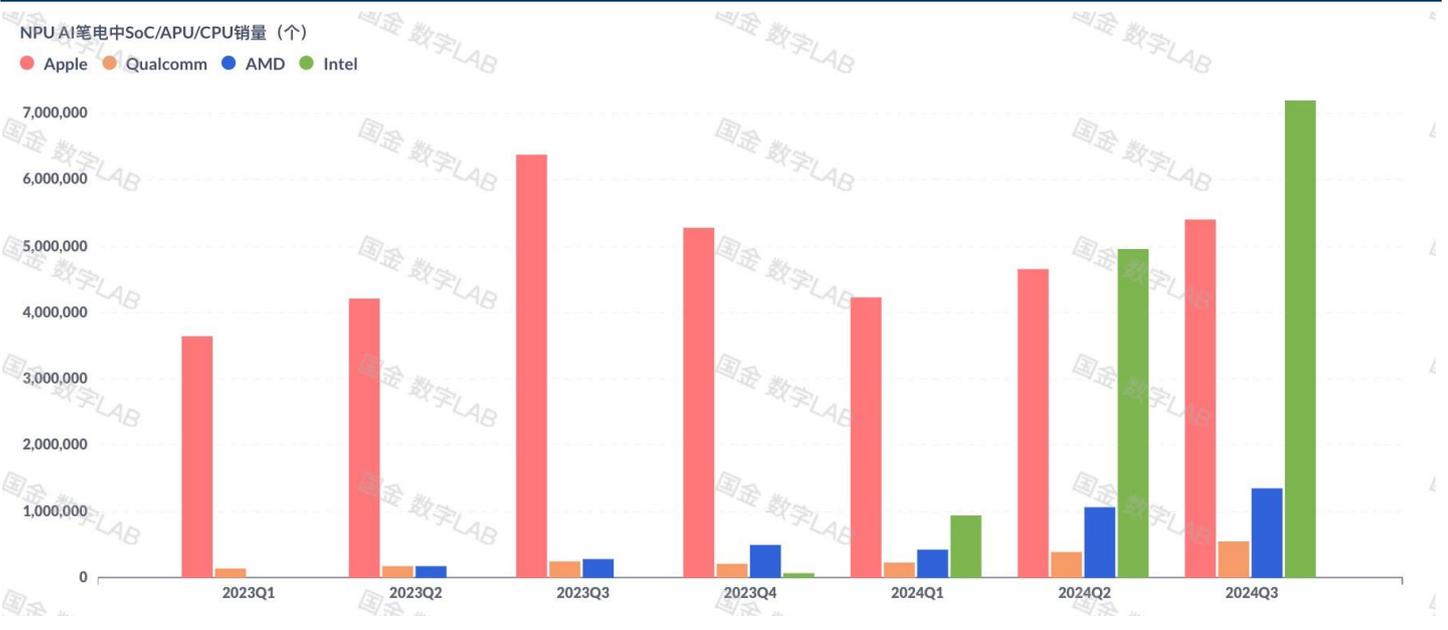


AI PC 销量

英特尔在 24Q3AI 笔电中表现良好

2024 年三季度，全球 AI 笔电销量约为 1450 万台，其中英特尔核 AI 笔电销量约为 720 万台，占比约 50%，苹果、AMD、高通核的 AI 笔电销量分别为 540 万、134 万、55 万台，占比分别为 37%、9%、4%。我们认为未来 AI 笔电 SoC/CPU 的竞争将会更为激烈。

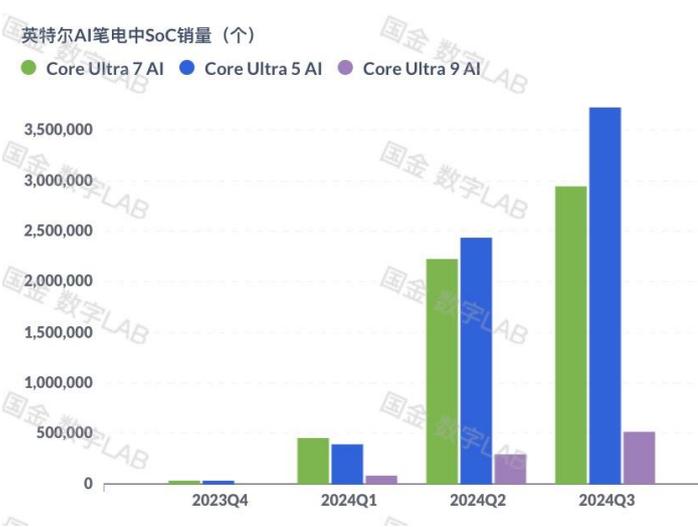
图表7: AI 笔电 SoC/CPU/APU 销量 (个)



来源: IDC、国金数字未来实验室、国金证券研究所

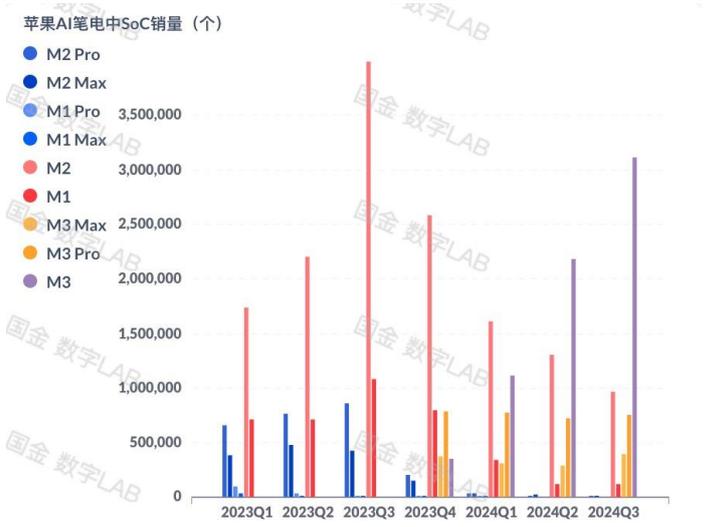
英特尔 Lunar Lake 系列芯片在 24Q2&24Q3 都有不错的销量，四季度英特尔发布了新一代 Ultra Lake 架构芯片，在功耗上有明显提升，我们认为英特尔核 AI 笔电四季度的销量将继续增长。苹果凭借 M3 芯片的强大性能维持销量的稳定，我们认为四季度 M4 核 AI PC 推出后，苹果销量将继续增长。AMD Ryzen 300 AI 系列在 8 月正式发布，在芯片性能表现优异的情况下我们认为 4 季度 Ryzen 300 AI 芯片的销量占比将会提高。高通由于 X Elite 在性能&价格等方面的劣势暂时销量较低，我们认为高通仍需时间进行改善。

图表8: 英特尔核 AI 笔电中 SoC 销量 (个)



来源: IDC、国金数字未来实验室、国金证券研究所

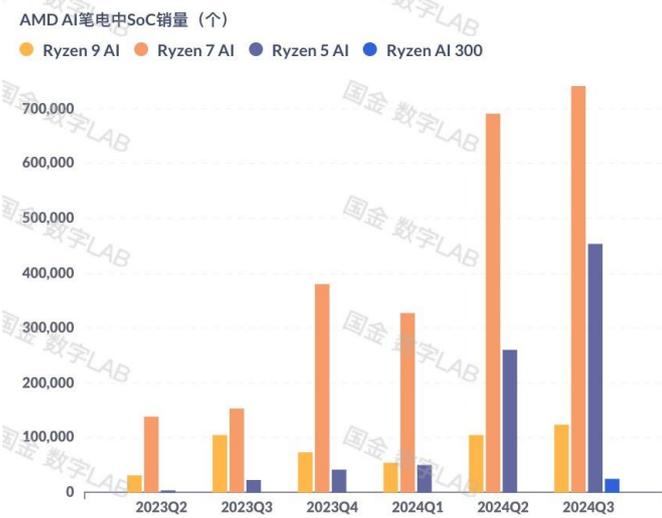
图表9: 苹果核 AI 笔电中 SoC 销量 (个)



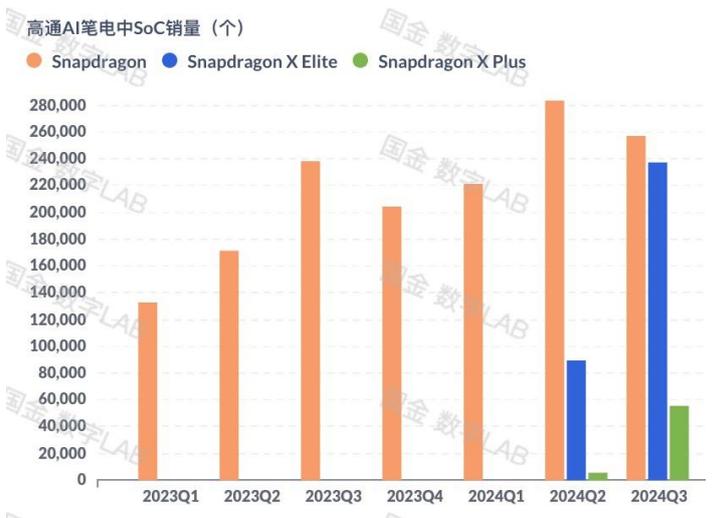
来源: IDC、国金数字未来实验室、国金证券研究所



图表10: AMD 核 AI 笔电中 SoC 销量 (个)



图表11: 高通核 AI 笔电中 SoC 销量 (个)



来源: IDC、国金数字未来实验室、国金证券研究所

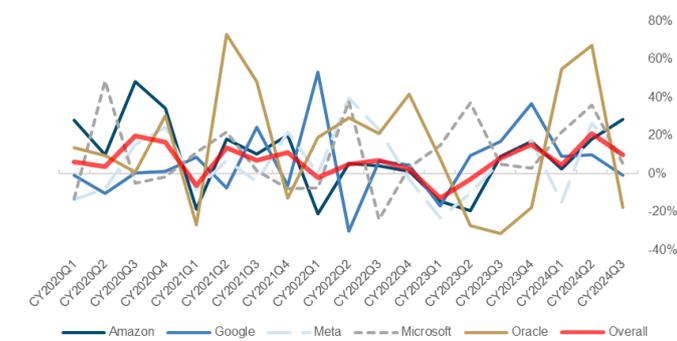
来源: IDC、国金数字未来实验室、国金证券研究所

财报前瞻

英伟达: 关注 Blackwell 首个交付季度表现

英伟达(NVIDIA)将于美国东部时间下周三盘后发布 2025 财年第三季度(FY25Q3)财报。值得关注的是, FY25Q3 已有少量 Blackwell 产品出货, 因此我们认为应重点关注本季度 Blackwell 产品的出货情况, 以及管理层对后续产品迭代的相关描述, 同时, 需要关注 Blackwell 系列产品的研发对公司毛利率的影响。

图表12: 云厂资本开支环比增速



来源: Reuters、国金证券研究所

图表13: 云厂资本开支同比增速

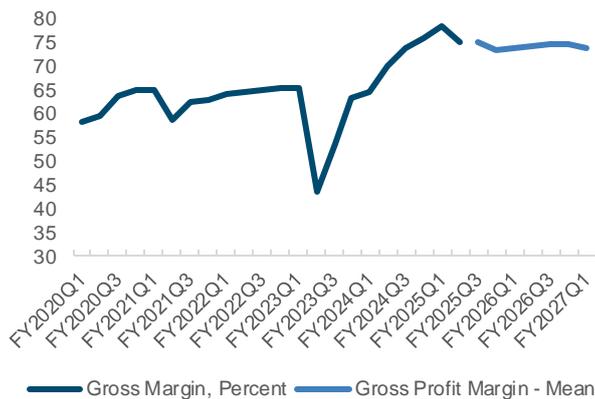


来源: Reuters、国金证券研究所

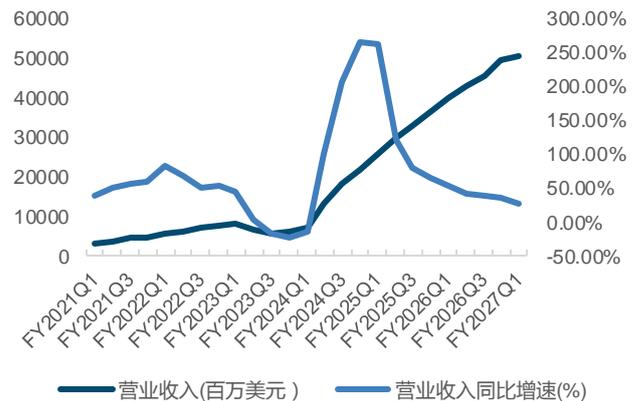
从需求端来看, 除亚马逊外, 其他主要云服务厂商的资本开支增速在上季度均有所放缓。这一趋势可能是受上游产品交付周期的影响, 而非采购意愿的收缩。例如, 微软在其财报电话会议中提到, Azure 本季度的收入增长受到供应链限制, 可能与 NVIDIA Blackwell GPU 的延迟交付有关。同时, 亚马逊的资本开支环比增速出现上升, 可能与其加速投入自研芯片有关。回顾此前几个季度, 资本开支环比增速本身存在一定波动, 因此此次增速放缓未必表明云服务厂商对 GPU 的需求减弱。展望未来, Meta 已将 2024 年资本开支下限从 370 亿美元上调至 380 亿美元, 并表示未来将持续加大投入; 亚马逊预计 2024 年全年资本开支将达到 750 亿美元。虽然其他厂商未公布具体的展望数据, 但均表示资本开支将继续增长。我们认为在 Blackwell 代际无需担忧下游需求不足的问题, 主要矛盾仍旧是关注公司产能和后续产品迭代节奏是否放缓。



图表14: 市场预期 FY25Q3 毛利率将有所下滑



图表15: 市场预期 FY25Q3 营业收入增速将低于 100%



来源: Reuters、国金证券研究所

来源: Reuters、国金证券研究所

根据路透社当前的一致预期,本季度毛利率为 75.01%,且未来几个季度不会有显著回升。这表明市场认为后续产品迭代所需的难度和研发投入将高于之前的世代,尤其是在摩尔定律放缓的背景下,这与我们此前的观点一致。

路透社一致预期营业收入为 329.14 亿美元,隐含的营业收入同比增速约为 80%。本世代产品采用了更大的芯片尺寸 (die size), 尽管 CoWoS 产能有所扩张,但单位封装中介层所能产出的 GPU 数量有所下降。我们认为这将是一个长期趋势,叠加此前高基数的影响,总体上产能所能支撑的营业收入增速下滑是在所难免的,市场当前的预期已反映出这一共识。

此外,我们认为市场尚未充分预期到下一代产品 Rubin 可能的延迟或性能提升不及预期的可能性。本世代产品 Blackwell 的延期已经在一定程度上显示出技术迭代速度的放缓,因此对下一代产品 Rubin 的预期不宜过高。建议密切关注本季度财报中管理层对未来代际产品的相关描述。

风险提示

1. 芯片制程发展与良率不及预期: 半导体工艺的发展面临诸多挑战, 主要包括技术瓶颈、良率提升难度、研发成本高企以及供应链不确定性等问题。随着工艺节点微缩变得愈发复杂, 先进制程的实现难度和成本不断攀升, 可能导致量产延迟, 甚至影响产品性能和成本控制。此外, 地缘政治风险和出口管制可能扰乱供应链, 进一步拖累产能扩张。
2. 中美科技领域政策恶化: 中美在 AI 领域竞争激烈, 美国限制先进芯片和半导体对中国的出口, 随着竞争的加剧, 未来可能会推出更严格的限制政策, 限制国内 AI 模型的发展。
3. 智能手机销量不及预期: 智能手机销量与产品本身质量关系紧密, 若产品本身有缺陷则智能手机销量可能收到影响。同时宏观经济变化也有可能导致消费者消费意愿发生变化从而影响智能手机销量。



特别声明:

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于C3级(含C3级)的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海	北京	深圳
电话: 021-80234211	电话: 010-85950438	电话: 0755-86695353
邮箱: researchsh@gjzq.com.cn	邮箱: researchbj@gjzq.com.cn	邮箱: researchsz@gjzq.com.cn
邮编: 201204	邮编: 100005	邮编: 518000
地址: 上海浦东新区芳甸路 1088 号 紫竹国际大厦 5 楼	地址: 北京市东城区建国内大街 26 号 新闻大厦 8 层南侧	地址: 深圳市福田区金田路 2028 号皇岗商务中心 18 楼 1806



【小程序】
国金证券研究服务



【公众号】
国金证券研究