

内容目录

1. 行业控费效果整体显著，但毛利率下降拖累了整体盈利能力	5
1.1. 行业收入增速呈放缓迹象	5
1.2. 减员控费效果明显，但是毛利率下滑成为拖累业绩的最主要因素	6
2. AI：2025 年推理端投入加快，软件应用端 AI 能力加快释放	6
2.1. AI 对行业的改变	7
2.1.1. 加速软件 SaaS 化趋势	7
2.1.2. AI 对云业务的促进作用明显	8
2.2. 模型发展的新趋势：推理端加快投入，软件应用端 AI 能力加快释放	9
2.3. 各个产业链环节的收入增长和盈利情况对比	11
3. 算力：国内需释放进度整体落后于海外，液冷需求释放值得期待	15
3.1. 英伟达 Blackwell 芯片量产延迟，但需求端仍然强劲	15
3.2. 海内外互联网厂商的资本投入创新高	17
3.3. 智算建设快于政策规划，AI 服务器市场长期还将保持强劲增长势头	19
3.4. 液冷服务器：未来 5 年预计 50% 的年复合增速，25 年运营商有望规模化应用	20
4. 国产化：纯血鸿蒙发布，芯片国产化加速	22
4.1. 美国禁令限制逐年深入	22
4.2. 运营商集采不断提高国产化比例，国产芯片供应商队伍持续壮大	24
4.2.1. 电信集采	24
4.2.2. 移动集采	24
4.2.3. 国产服务器供应商队伍持续壮大	25
4.3. 纯血鸿蒙发布，国产操作系统阵营取得新突破	26
4.4. 芯片国产化进程持续深入，利好 EDA 等上游产业链环节	27
4.5. 基础软件国产化率持续提升	28
4.5.1. 数据库：国产厂商约占 2/3 的份额	28
4.5.2. 操作系统：欧拉在服务器操作系统市占率将达 50%	30
5. 投资策略	31
5.1. 行情回顾：总体略好与沪深 300，领涨概念呈现强者恒强特征	31
5.2. 估值：行业的估值位于历史均值之间	32
5.3. 行业观点与投资建议	33
6. 风险提示	34

图表目录

图 1：2018-2024 年我国软件业务收入及增速（月度累计值）	5
图 2：2019-2024 年前三季度计算机行业收入增速对比	5
图 3：19-24 年前三季度计算机行业扣非净利润增速对比	5
图 4：2018-2023 年计算机行业员工数量及增速	6
图 5：2019-2024 年前三季度计算机行业三大费用增速	6
图 6：2019-2024 年前三季度计算机行业盈利能力指标	6
图 7：2019-2024 年前三季度计算机行业毛利额增速对比	6
图 8：2021H1-2024H1 我国公用云三大细分市场的增速	7
图 9：金山办公 AI 产品家族	7
图 10：2021Q1-2024Q3 用友网络小微企业业务收入及增速	8
图 11：2014-2024 年海外云计算厂商云业务相关收入及增速（亿美元）	9
图 11：推理时间计算	10

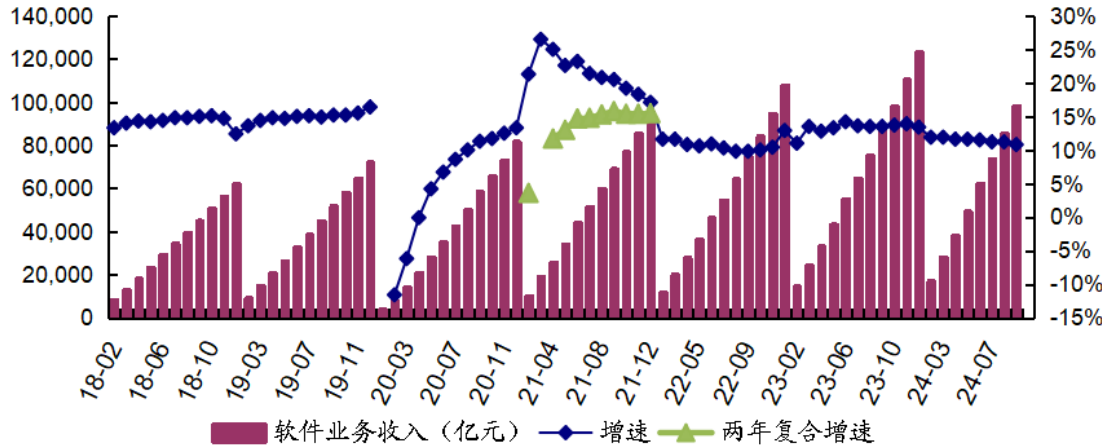
图 12: 2024 年 10 月 AI 应用榜单 (中国)	11
图 13: 2024 年 10 月 AI 应用榜单 (海外)	11
图 14: 2020-2024 年英伟达单季度收入规模 (万美元)	12
图 15: 2020-2024 年英伟达单季度净利率 (%)	12
图 16: 2020-2024 年浪潮信息单季度收入规模 (万元)	12
图 17: 2020-2024 年浪潮信息单季度净利率 (%)	12
图 18: 2020-2024 年万国数据单季度收入规模 (万美元)	13
图 19: 2020-2024 年万国数据单季度净利率 (%)	13
图 20: 2020-2024 年润泽科技单季度收入规模 (万元)	13
图 21: 2020-2024 年润泽科技单季度净利率 (%)	13
图 22: 2020-2024 年阿里云单季度收入规模及盈利能力 (亿元)	13
图 23: OpenAI 的收入预测	14
图 24: 2021Q1-2024Q3 三大芯片厂商数据中心业务收入及增速 (亿美元)	16
图 25: 20Q1-24Q3 5 大科技厂商资本投入 (百万美元)	17
图 26: 2020Q1-2024Q3BAT 的资本开支 (亿元)	17
图 27: 2022.6-2025E 我国算力规模及结构 (EFLOPS)	19
图 28: 2018H1-2024H1 我国 AI 服务器市场规模及增速	19
图 29: 2024-2028 中国 AI 服务器市场预测 (百万美元)	19
图 30: 2024H1 我国液冷服务器市场厂商份额	21
图 31: 2024-2028 年我国液冷服务器市场预测 (百万元)	21
图 32: 2022 年我国液冷服务器市场厂商份额	21
图 33: 2023 年我国液冷服务器市场厂商份额	21
图 34: 液冷技术与数据中心能耗指标关系图	22
图 35: 液冷技术与数据中心能耗指标关系图	22
图 36: 2022Q1-2024Q3 英伟达单季度收入的区域分布 (亿美元)	23
图 37: 2020-2024 年电信服务器集采订单中服务器采购金额及国产化率情况 (亿元)	24
图 38: 2022Q2-2024Q1 全球智能手机操作系统份额	27
图 39: 2022Q2-2024Q1 中国智能手机操作系统份额	27
图 40: 2021-2024 年我国集成电路对进口依赖度和出口占比	27
图 41: 2023H1-2024H1 我国 AI 服务器芯片国产化占比	28
图 42: 2018-2024 年 EDA 企业国内业务收入规模及增速 (万元)	28
图 43: 2019 本地部署模式关系型数据库市场份额	30
图 44: 2024H1 本地部署模式关系型数据库市场份额	30
图 45: 2019 公有云模式关系型数据库市场份额	30
图 46: 2024H1 公有云模式关系型数据库市场份额	30
图 47: 2020-2024 年欧拉系 OS 占中国服务器操作系统新增市场份额	31
表 1: 国内主要大模型应用状况	14
表 2: 英伟达、AMD 人工智能芯片参数对比	15
表 3: 三大运营商服务器集采情况及国产化比例	25
表 4: 三大运营商服务器集采情况及国产化比例	25
表 5: 重点推荐公司的估值情况 (截止 2024.11.28)	33

1. 行业控费效果整体显著，但毛利率下降拖累了整体盈利能力

1.1. 行业收入增速呈放缓迹象

相比于 2023 年来看，我国软件产业增速在 2024 年呈现持续回落趋势。根据工信部数据，2024 年 1-9 月软件业务收入 9.83 万亿元，同比增长 10.8%，较 2023 年 13.4% 的增速回落了 2.6 PCT，较 1-8 月增速下滑了 0.4 PCT。

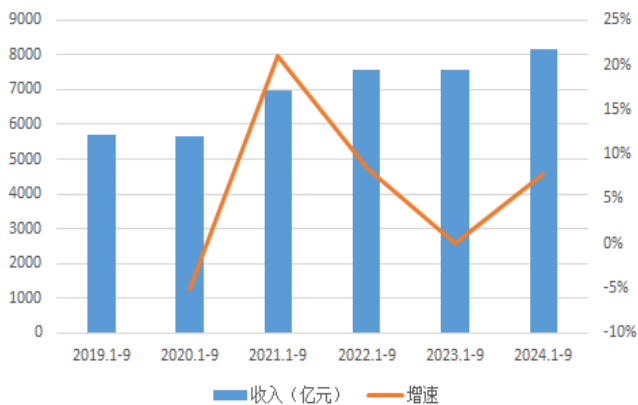
图 1：2018-2024 年我国软件业务收入及增速（月度累计值）



资料来源：工信部，中原证券研究所

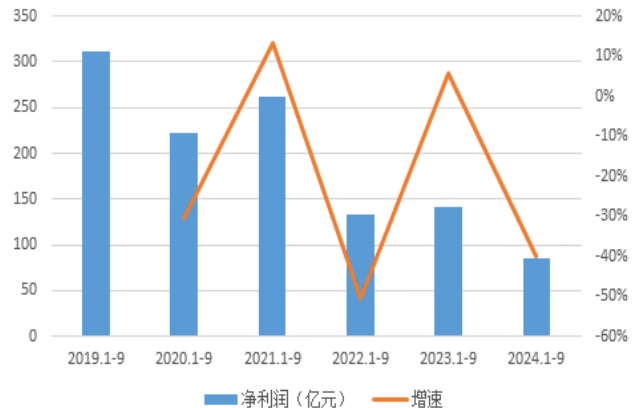
财报数据来看，行业收入持续保持增长势头，但是利润下滑明显。从中信计算机行业数据来看，2024 年前三季度，行业总计收入 1.25 万亿元。考虑到工业富联占比过高（约占比行业总收入的 35%，以下行业财报数据中均做剔除处理），将其剔除后，行业总计收入 8170.94 亿元，同比增长 7.75%，较上年同期提升了 7.70PCT。行业前三季度行业净利润 84.54 亿元，同比下滑了 39.87%，较上年同期下降了 45.57 PCT；扣非净利润 15.21 亿元，同比下滑了 63.84%，较上年同期下降了 10.56 PCT，已经连续三年呈现出下降的趋势。

图 2：2019-2024 年前三季度计算机行业收入增速对比



资料来源：iFind，上市公司公告，中原证券研究所（除工业富联）

图 3：19-24 年前三季度计算机行业扣非净利润增速对比



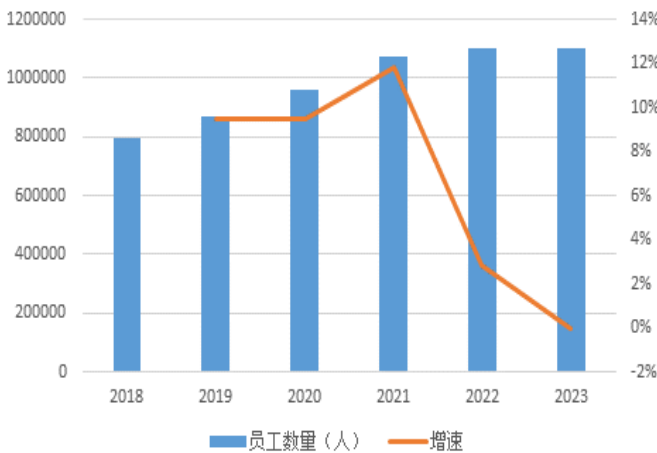
资料来源：iFind，上市公司公告，中原证券研究所（除工业富联）

1.2. 减员控费效果明显，但是毛利率下滑成为拖累业绩的最主要因素

2023年，计算机行业员工数量总计110.22万人，同比下滑了0.05%，2024年继续呈现缩减趋势。因而我们看到2024年前三季度，行业研发费用、管理费用、销售费用同比分别增长了-2.24%、1.74%、-1.49%，三大费用总计下滑了1.02%。

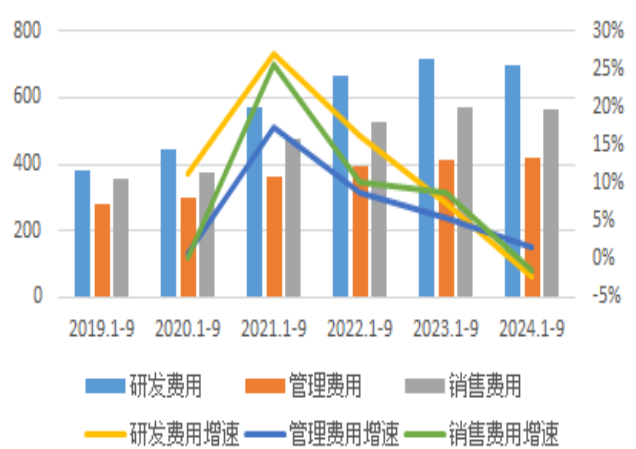
考虑到2024年人员缩减会带来解除劳动合同相关的赔付费用支出，再叠加人员缩减的滞后效应，我们预计2025年行业三大费用还将呈现进一步缩减的趋势。

图 4：2018-2023 年计算机行业员工数量及增速



资料来源：Wind，上市公司公告，中原证券研究所（除工业富联）

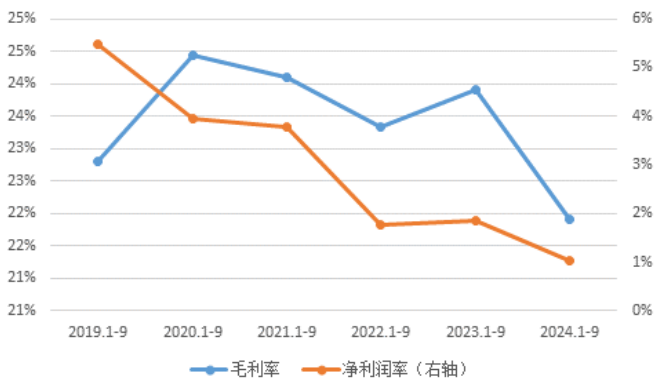
图 5：2019-2024 年前三季度计算机行业三大费用增速



资料来源：Wind，上市公司公告，中原证券研究所（除工业富联）

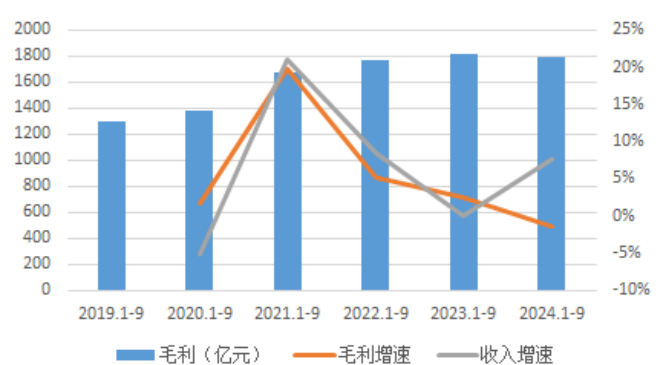
2024年前三季度，行业毛利率下降了2.01PCT，在收入增长7.75%的背后，公司毛利额下滑1.31PCT，成为拖累当期业绩最重要的因素。

图 6：2019-2024 年前三季度计算机行业盈利能力指标



资料来源：iFind，上市公司公告，中原证券研究所（除工业富联）

图 7：2019-2024 年前三季度计算机行业毛利额增速对比



资料来源：iFind，上市公司公告，中原证券研究所（除工业富联）

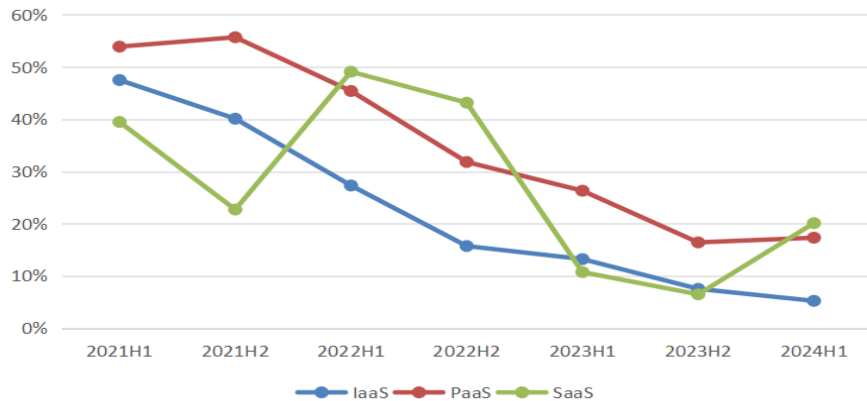
2. AI：2025 年推理端投入加快，软件应用端 AI 能力加快释放

2.1. AI 对行业的改变

2.1.1. 加速软件 SaaS 化趋势

根据 IDC 数据进行测算，2024 上半年我国 SaaS 市场增速实现了见底回升，从 2023H2 的 6.5% 回升到了 20.1%。我们认为软件 SaaS 化加速，一方面源于企业预算和降本增效的考虑，另一方面也是企业推动 AI 创新的需要。

图 8：2021H1-2024H1 我国公用云三大细分市场的增速



资料来源：IDC，中原证券研究所

从国内头部 SaaS 企业的业务进展来看，2024 年 SaaS 化加速趋势明显。

金山办公的 AI 产品除了提升 C 端会员付费比例和人均付费金额以外，在 B 端也加速了机构订阅业务的 SaaS 化。公司前三季度 B 端订阅收入中，地方国企、民企的收入占比快速提升，预期 2024 年中的 SaaS 收入有望实现数倍增长。

图 9：金山办公 AI 产品家族

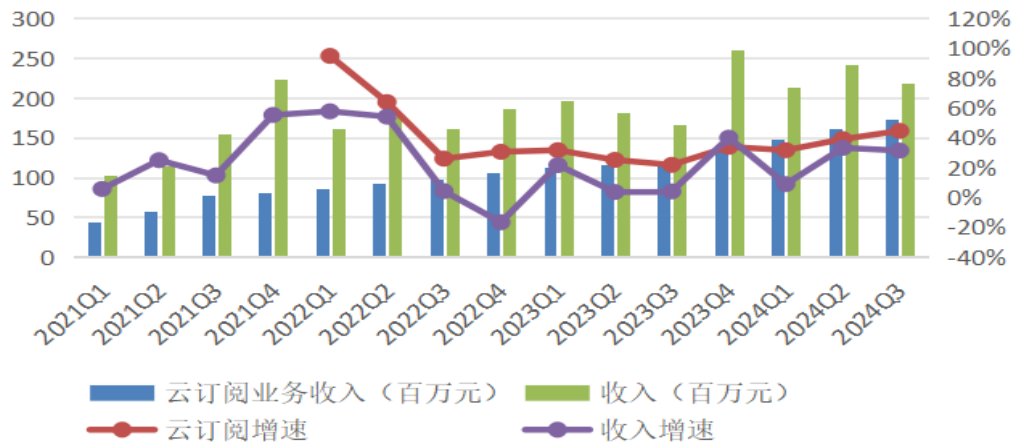


资料来源：金山办公公告，中原证券研究所

用友的小微企业业务作为其率先完成云转型的业务，收入增速在 2022Q4 见底以后总体呈现回升态势，其中云订阅业务比例也在逐季提升。这期间，小微企业业务加快 AI 技术在产

品创新及公司运营中的应用，实施全面效益化经营，持续改善盈利能力。此外，用友的其他软件产品 2024 年也展开了全面订阅化，7 月推出了面向一般大型企业的 BIP 超级版，前三季度中型企业客户订阅收入同比增长 61.8%，核心产品 YonSuite 续费率达到 90.3%。

图 10: 2021Q1-2024Q3 用友网络小微企业业务收入及增速



资料来源：用友网络公告，中原证券研究所

在加快 SaaS 化转型的同时，企业可以降低定制化所需要的额外成本，规模效应逐步增强。因而短期虽然会带来收入确认节奏的变化，长期进入到年收入更为稳定，头部优势持续增强的新发展格局中。

2.1.2. AI 对云业务的促进作用明显

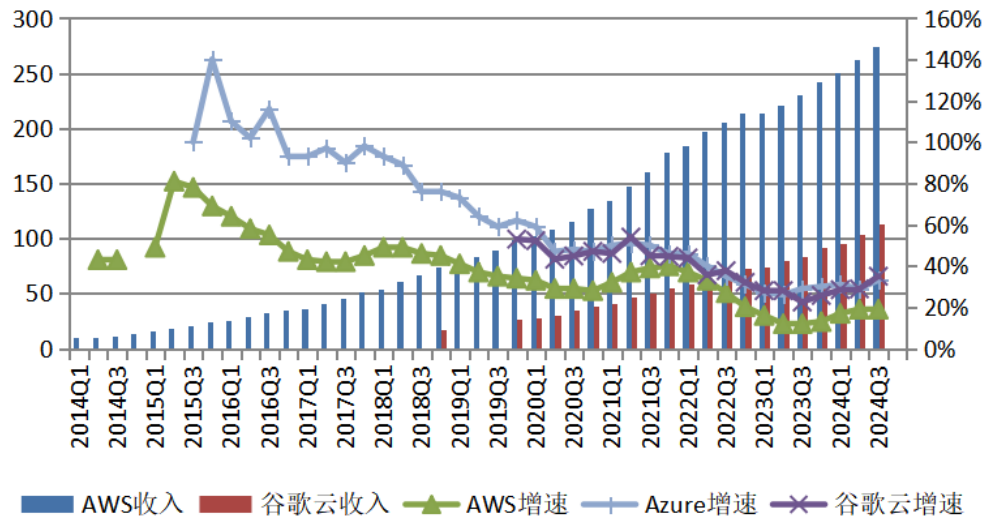
AI 业务加持下，海外云计算头部厂商加速增长：

(1) **AWS**：3 季度亚马逊云计算业务 AWS 实现收入 274.5 亿美元，同比增长 19%，高于上年同期 12% 的增速水平，其中来自生成式 AI 的收入达到数十亿美元，继续实现了三位数的同比增速，同时 AWS 的营运利润率达到历史新高。

(2) **Azure**：微软 Azure 和其他云服务营收增长 33%，较上季度提升 4 PCT，其中有 12 个百分点的增长来自 AI。

(3) **谷歌云**：受到 AI 业务的驱动，谷歌云业务收入 113.53 亿美元，同比增长 35%，较上季度提升了 6 PCT，旗下全栈 AI 产品已经实现大规模运营，被十亿用户创造出良性循环，Gemini AP 调用量增加近 14 倍，同时也在努力降低 AI 的使用成本。

图 11：2014-2024 年海外云计算厂商云业务相关收入及增速（亿美元）



资料来源：上市公司财报，中原证券研究所

国内方面：

(1) 阿里云：由于 AI 相关产品收入连续 5 个季度实现三位数的同比增长，阿里云的公共云业务实现了两位数增长，并带动阿里云整体收入增长 7% 达到 296.1 亿元，连续 4 个季度实现收入增速的增长。同时，阿里云 Q3 经调整 EBITA 利润达到 26.61 亿元，同比继续大幅增长了 89%。AI 对公司公有云业务带动作用下，阿里云一方面积极加大在 AI 基础设施方面的投入，一方面逐步降低了非公有云收入和提高收入质量。

(2) 金山云：2024Q3 金山云实现收入 26.91 亿元，同比增长 16%，连续 3 个季度收入增速增长，其中人工智能业务持续扩展至 3.65 亿元，同时受益于电动车、大模型和 WPS AI 等领域的商机，金山云来自小米及金山生态系统的收入也大幅增长了 36%，EBITA 利润率快速上升到 9.8%。

通过以上厂商数据可以看到，AI 在给云厂商业务加速的同时，也在助力云厂商的业务转型，从而提升了收入质量，改善了整体盈利水平。

2.2. 模型发展的新趋势：推理端加快投入，软件应用端 AI 能力加快释放

纵观本轮人工智能在大语言模型方向上取得的成就，就是因为发现了 Transformer 符合 Scaling Laws，通过增加模型参数、训练数据和算力，保持系统能力的增长，甚至在规模达到一定程度，模型能力会得到指数级的提升，也就是涌现。

而随着算力规模指数级的增长，训练数据枯竭，合成数据也未能取得理想的效果，在预训练方向的 Scaling Laws 面临发展瓶颈。11 月 10 日，根据 The Information 报道，OpenAI 的下一代旗舰模型 Orion 可能不会像前面几代产品那样带来巨大的飞跃。Orion 性能虽然超过了 OpenAI 现有的模型，但是面临了新训练数据匮乏的瓶颈。OpenAI 意图通过合成数据进行

Orion 的训练，但是合成数据可能会导致 Orion 在某些方面与那些旧模型相似。

作为大模型发展的风向标，OpenAI 2024 年并没有推出预期中的 GPT-5，而是推出了 Sora、GPT-4o 和 o1，前者两者分别聚焦文生视频和多模态交互能力，o1 提升了大模型解决科学、编码、数学和类似领域的复杂问题。此前科学界对于 LLM 的技术路线能否在提高语言能力的同时，高效提高逻辑能力和推理能力存在较大质疑，因而 o1 对于 LLM 实现 AGI 具有重要的意义，是 OpenAI 在模型能力提升上更为重要的突破。

表 1: OpenAI 大模型产品

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	1.8 万亿	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		速度是 GPT4 的 5 倍，价格降低 2/3
Sora	2024.2					首个文生视频模型
GPT-4o	2024.5					具有了强大的多模态交互能力 对用户免费,API 速度较 GPT-4 Turbo 提高 2 倍，价格降低 50%
o1	2024.9					更擅长解决科学、编码、数学和类似领域的复杂问题

资料来源：OpenAI，中原证券研究所

o1 模型将更多的算力投入到了推理阶段，采用了“推理时间技术”，当模型面临逻辑推理类复杂任务时，能够实时生成和评估多种解答路径，从而选择最佳解决方案，为模型开拓了一个新的提升路径，同时也将带来 AI 对于推理端算力需求提升的趋势，利好国产芯片厂商。

图 11: 推理时间计算

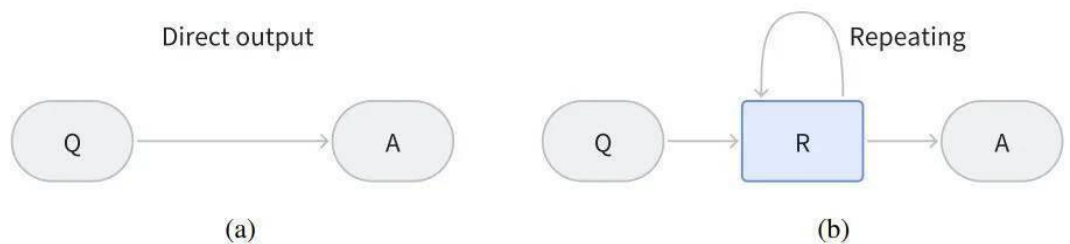


Figure 1: Inference-time computation. (a) An autoregressive LLM directly generate an answer (A) by conditioning on the given question (Q). (b) The concept of chain of thought, or step-by-step thinking, involves incorporating intermediate reasoning steps (R) before arriving at the final answer (A). These repeated operations allow for 1) revisiting and revising prior outputs, 2) progressing to subsequent reasoning stages, and 3) exploring multiple reasoning paths or trajectories.

资料来源：机器之心，中原证券研究所

随着 o1 模型对大模型逻辑和推理能力的补足，AI Agent 也步入了发展的快车道，AI 在应用软件层的发展潜力亟待释放。10 月 22 日，微软宣布在 Dynamics 365 中集成 10 个自主

AI Agent, 可以帮助企业自动执行客服、销售、财务、仓储等业务, 开创全新的智能自动化工作模式。

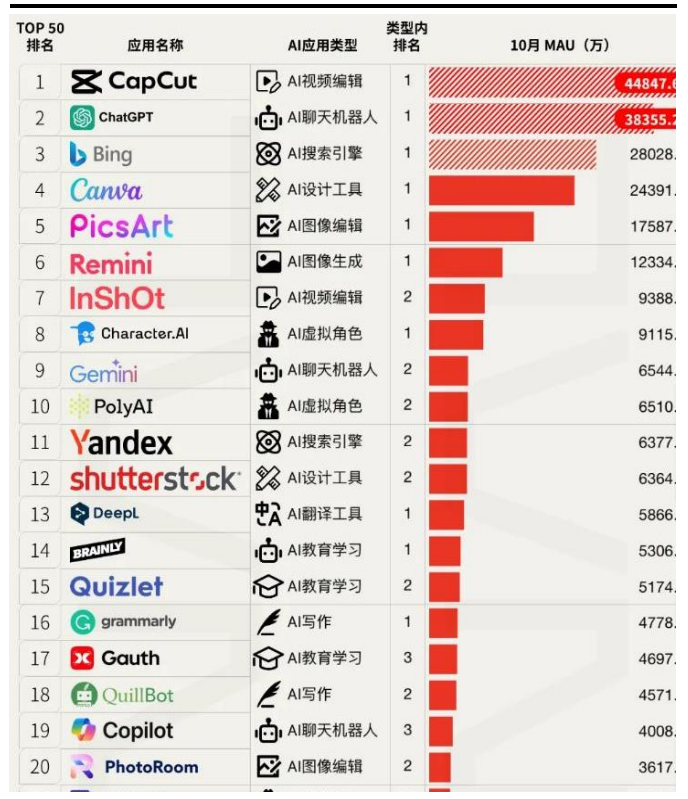
此外, 我们看到 PC、手机端等越来越多的硬件端加快了 AI 的渗透, AI 功能的嵌入也成为软件端不可或缺的组成部分, AI 应用即将从能用步入到好用的阶段。

图 12: 2024 年 10 月 AI 应用榜单 (中国)



资料来源: 奇异因子, 中原证券研究所

图 13: 2024 年 10 月 AI 应用榜单 (海外)



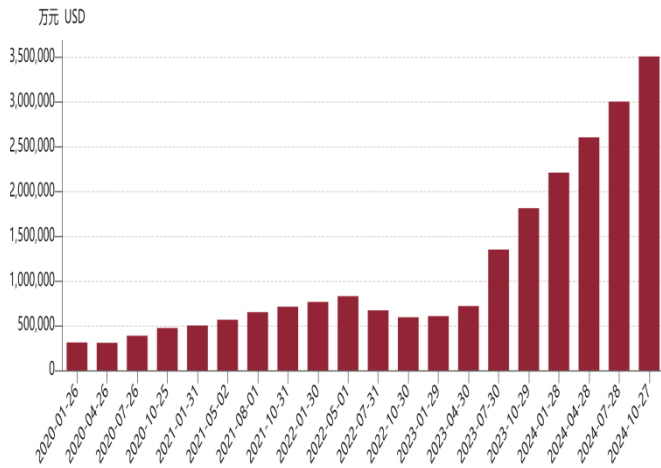
资料来源: 奇异因子, 中原证券研究所

2.3. 各个产业链环节的收入增长和盈利情况对比

从各产业链环节的代表性厂商的业绩来看:

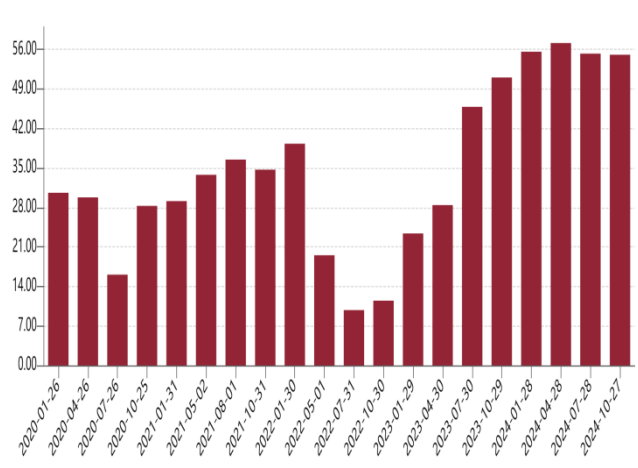
(1) GPU 芯片: 英伟达净利率从此前的 30%左右上升到最新一个季度的 55%, 收入从 2022Q3 的 53.3 亿美元增长到 2024Q3 的 350.82 亿元。

图 14: 2020-2024 年英伟达单季度收入规模 (万美元)



资料来源: Wind, 中原证券研究所

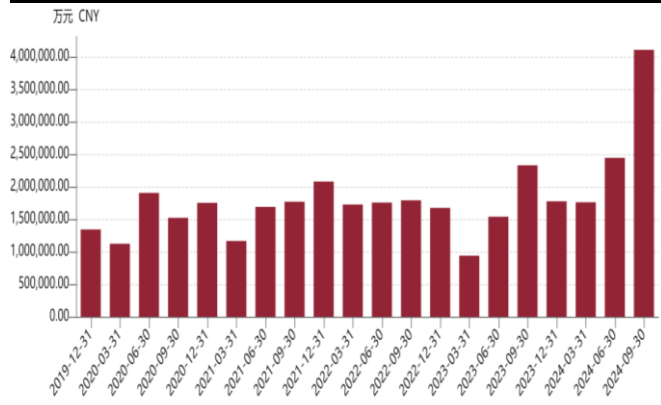
图 15: 2020-2024 年英伟达单季度净利率 (%)



资料来源: Wind, 中原证券研究所

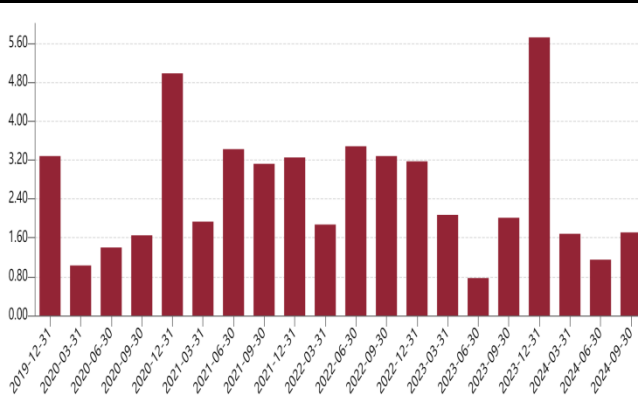
(2) 服务器: 作为国内最大的服务器厂商, 浪潮收入规模从 2022Q3 的 167.59 亿元增长到 2024Q3 的 410.62 亿元, 创出了历史新高, 但是在收入快速增长的同时, 公司收入结构变化导致毛利率下降, 盈利能力也处于相对较低水平, 2024Q1 净利率为 1.71%。

图 16: 2020-2024 年浪潮信息单季度收入规模 (万元)



资料来源: Wind, 中原证券研究所

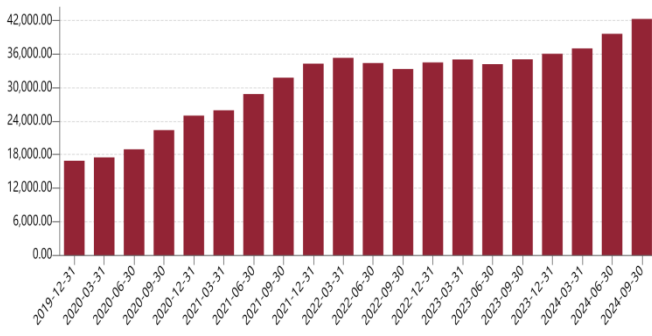
图 17: 2020-2024 年浪潮信息单季度净利率 (%)



资料来源: Wind, 中原证券研究所

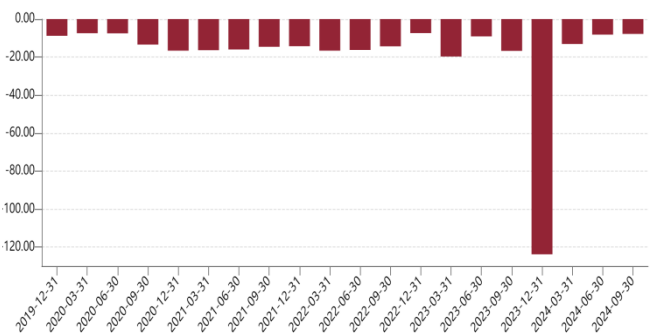
(3) 数据中心: 作为国内最大的第三方 IDC 厂商, 万国数据收入规模在 2023 年几乎保持了稳定, 2024 年开始恢复增长, 2022Q3 收入规模 3.33 亿美元, 2024Q3 增长到 4.23 亿美元。在盈利能力方面, 万国数据依然处于亏损状态, 但盈利能力呈现改善状态, 2024Q3 净利率-7.79%, 2022Q3 为-14.35%。

图 18: 2020-2024 年万国数据单季度收入规模 (万美元)



资料来源: Wind, 中原证券研究所

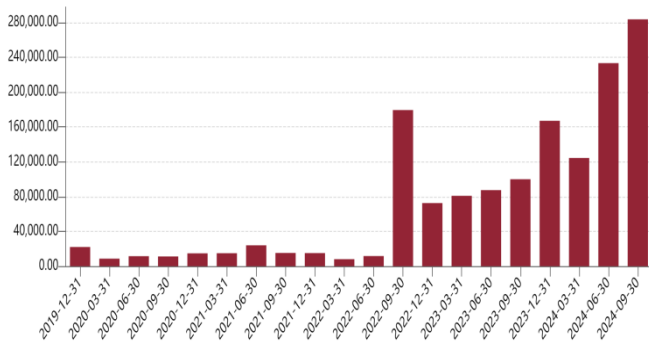
图 19: 2020-2024 年万国数据单季度净利率 (%)



资料来源: Wind, 中原证券研究所

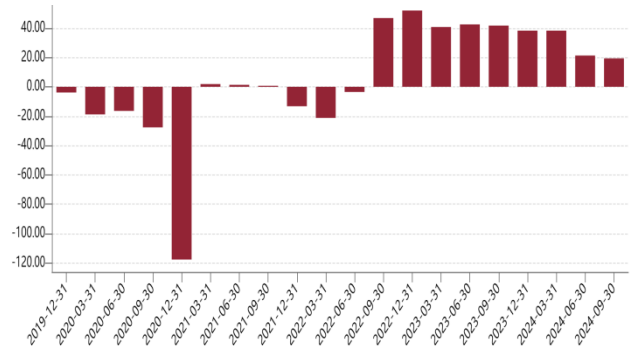
而同样作为数据中心厂商, 润泽科技作为字节的供应商, 润泽在 2022 年借壳上市, 2022Q4 收入 7.25 亿元, 快速增长到 2024Q3 的 28.35 亿元, 但与此同时公司毛利率下滑, 导致净利率从 2022Q4 的 52.01% 下滑到了 2024Q3 的 19.36%, 这其中原因包括了 AIDC 毛利率相对较低和新建数据中心带来的折旧摊销费用增长等因素。

图 20: 2020-2024 年润泽科技单季度收入规模 (万元)



资料来源: Wind, 中原证券研究所

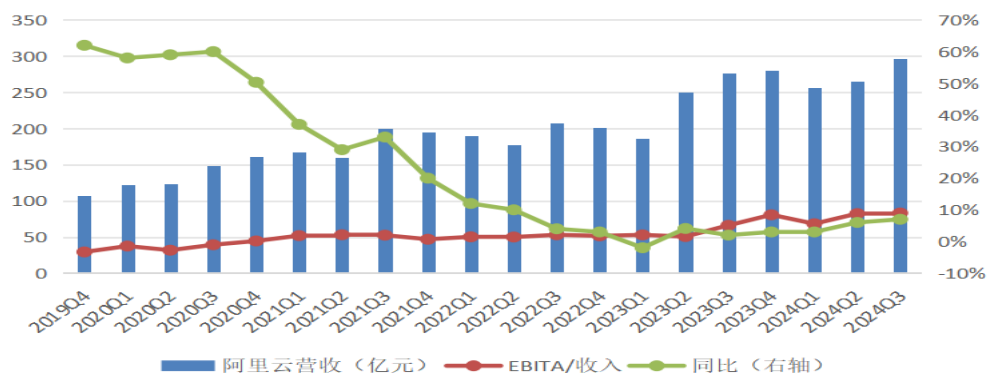
图 21: 2020-2024 年润泽科技单季度净利率 (%)



资料来源: Wind, 中原证券研究所

(3) 云计算: 作为国内云计算行业龙头, 近两年来阿里云收入仅保持了个位数的增长, 近 4 个季度都呈现增速提升趋势, 同时经调整 EBITA 与收入比值从 2022Q3 的 2% 提升到了 2024Q3 的 9%。

图 22: 2020-2024 年阿里云单季度收入规模及盈利能力 (亿元)



资料来源: 阿里财报, 中原证券研究所

(4) 大模型厂商：根据财联社消息，据媒体 The Information 获得的 OpenAI 公司财务文件，预计从 2023 年到 2028 年，OpenAI 将蒙受 440 亿美元的损失，其中今年将损失 50 亿美元，到 2026 年年度损失可能高达 140 亿美元，到 2029 年 OpenAI 或才能扭亏为盈。

图 23: OpenAI 的收入预测



资料来源：The Information，中原证券研究所

从国内大模型厂商来看，商汤虽然生成式 AI 收入快速增长，由于传统业务收入的减少，收入从 2022H1 的 14.15 亿元增长到 2024H1 的 17.41 亿元，增幅有限，净利润从 2022H1 的 -31.58 亿元减少到 -24.57 亿元，仍有较大的盈利压力；科大讯飞收入从 2022Q3 的 47.72 亿元增长到 2024Q3 的 55.25 亿元，受到 G 端业务下滑的影响，增幅有限，但是已经在加大大模型研发投入和布局的同时在 2024Q3 实现了利润转正。

表 1: 国内主要大模型应用状况

公司	近况
百度	文心一言的用户规模从去年 11 月的 7000 万增长至今年 11 月的 4.3 亿，一年增幅超 500%；飞桨文心开发者数量已达 1808 万，服务了 43 万家企业，创建了 101 万个模型；截至今年 11 月初，百度文心大模型的日均调用量已经超过 15 亿次，相较一年前首次披露的 5000 万次，增长约 30 倍
科大讯飞	“讯飞星火”APP 在安卓端的下载量超过 1.9 亿次，国内工具类通用大模型 APP 排名第一。2024 年前三季度新增开发者数量达 177.7 万，公司正与 41.5 万余家企业客户用星火创新应用体验。“讯飞星火”大模型成为央企国企首选，1-9 月份，科大讯飞的大模型相关项目的中标数量和金额在行业里都处于领先地位。
商汤	2024 年上半年，生成式 AI 业务实现 255.7% 的大幅增长，收入达到 10.51 亿元，占公司整体收入比重 60.4%，3000 多家行业头部企业使用商汤的大模型及 AIDC 服务，公司 AIDC 运营总算力突破 20000PetaFLOPS，较 2024 年 3 月底约为 12000PeteFLOPS，上线 GPU 数量达到 5.4 万张，持续提升万卡集群 GPU 资源利用率至最高 80% 以上。
阿里	截止 2024 年 11 月，全球开发者基于阿里“通义千问”二次开发的衍生模型突破 7.8 万个，超 800 万用户通过阿里 AI 模型社区“魔搭”下载模型、投入应用。
月之暗面	截止 10 月 Kimi 的每月使用人数已达到 3600 万。

资料来源：百度，36 氪，上市公司公告，中原证券研究所

从以上数据我们可以看到，目前大模型产的发展过程中，最收益的仍然是上游的芯片厂

商，实现了收入和盈利上的双重改善，服务器端增收迹象明显，但是由于互联网客户采购毛利低，净利率下滑明显，数据中心收入增长加快，但短期盈利能力或受到新建数据中心冲击，云计算也有 AI 业务加速和盈利改善迹象，大模型厂商目前仍然处于投入期。考虑到 2025 年大模型应用的进一步增长和国产 AI 算力的释放，我们认为产业下游盈利改善有望加速。

3. 算力：国内需释放进度整体落后于海外，液冷需求释放值得期待

3.1. 英伟达 Blackwell 芯片量产延迟，但需求端仍然强劲

英伟达在 2024 年 3 月发布了基于 Blackwell 新架构的 B200 和 GB200 芯片（由 2 个 B200 GPU 和 1 个 Grace CPU 组成的超级芯片），原本计划在 2024 年 3 季度量产，4 季度开始批量出货。而后在 8 月初，美国科技网站 The Information 报道，由于设计缺陷，B200 出货时间将被延迟至少 3 个月。Blackwell 封装是首款采用台积电 CoWoS-L 技术进行封装的大批量设计，面临设计和产能问题，导致英伟达向服务器 OEM 客户推出采用 CoWoS-S 封装技术的降规版 B200A，而后又调整为 B300A。

英伟达 B200 和 GB200 预计将于 2024Q4 和 2025Q1 开始陆续出货，B300 系列预计将在 2025Q2-Q3 推出，B300A 将于 2025 年中期进入量产阶段。在 2025 财年三季报发布期间，英伟达 CFO 表示 Blackwell 芯片已发货 13000 件。

表 2：英伟达、AMD 人工智能芯片参数对比

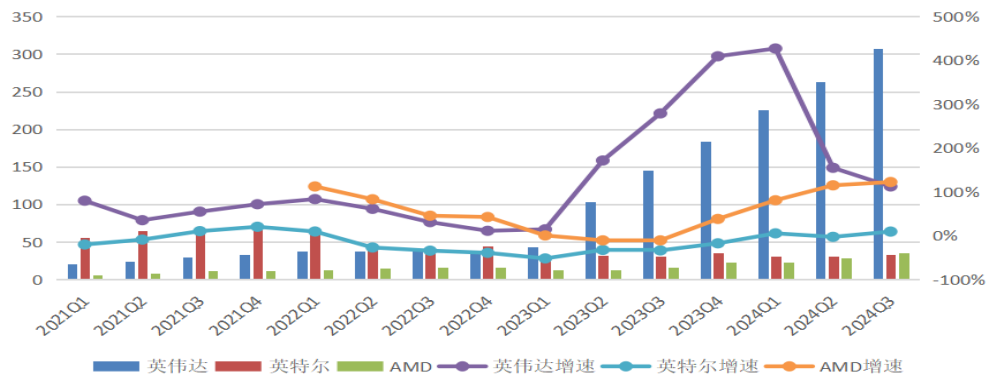
	H100 SXM	H200 SXM	B200	B300	MI300X	MI325X	MI355X
厂商	英伟达	英伟达	英伟达	英伟达	AMD	AMD	AMD
发布时间	2022.3	2023.11	2024.3	2025	2023.6	2024.6	2025
架构	Hopper	Hopper	Blackwell	Blackwell Ultra	CDNA 3	CDNA 3	CDNA 4
工艺	4nm	4nm	4nm		5nm	5nm	3nm
晶体管数量	800 亿	800 亿	2080 亿		1530 亿	1530 亿	
最大功率	700W	700W	1000W		750W	1000W	
显存容量	80GB	141GB	192GB	288GB	192GB	256GB	288GB
显存带宽	3.35TB/s	4.8TB/s	8TB/s		5.3TB/s	6TB/s	8TB/s
显存颗粒	HBM3	HBM3e	HBM3e 8hi	HBM3e 12hi	HBM3	HBM3e	HBM3e 12hi
传输速率	900GB/s	900GB/s	1800GB/s		896GB/s		
FP64	34 TFLOPS	34 TFLOPS					
FP64 Tensor Core	67 TFLOPS	67 TFLOPS					
FP32	67 TFLOPS	67 TFLOPS					
TF32					653.7 TFLOPS		

TF32 Tensor Core	989 TFLOPS	989 TFLOPS	2.2 PFLOPS				
BFLOAT16	989.4 TFLOPS				1307.4 TFLOPS		
BFLOAT16 Tensor Core	1979 TFLOPS	1979 TFLOPS					
FP16	989.4 TFLOPS				1307.4 TFLOPS	1.3 PFLOPS	2.3 PFLOPS
FP16 Tenor Core	1979 TFLOPS	1979 TFLOPS	4.5 PFLOPS				
FP8	1978.9 TFLOPS				2614.9 TFLOPS	2.6 PFLOPS	4.6 PFLOPS
FP8 Tensor Core	3958 TFLOPS	3958 TFLOPS	9 PFLOPS				
INT8	1978.92TOPS				2614.92 TOPS		
INT8 Tensor Core	3958 TOPS	3958 TOPS	9 PFLOPS				
FP6	-	-	9 FLOPS		-		9.2 PFLOPs
FP4	-	-	18 PFLOPS		-		9.2 PFLOPs

资料来源：英伟达，AMD，中原证券研究所

从目前 GPU 市场来看，英伟达依然处于主导地位。英伟达数据中心业务在 2024Q1 增速见顶，Q2、Q3 连续两个季度增速下滑，但是依然保持了翻倍增长。Blackwell 供给不及预期，但是市场需求仍然强劲，Hopper 出货仍在持续放量。2024Q4 和 2025Q1 市场对于 B200 和 GB200 的供货和放量仍然充满期待，但等待新品或将导致需求延迟释放。

图 24：2021Q1-2024Q3 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，Wind，中原证券研究所

但与此同时，我们也看到硬件端一些新的发展趋势：

(1) 当前 GPU 的算力提升放缓，内存提升加速。英伟达此前的产品有 4-5 倍的算力提升，而到了 Blackwell 实质性的提升只有 2.5 倍。大模型需求下，内存容量、内存带宽和互联带宽的重要性显著提升，算力重要性降低，因而降低了英伟达的护城河，但内存上的差距和禁令影响也加大了国产厂商赶超的进程。

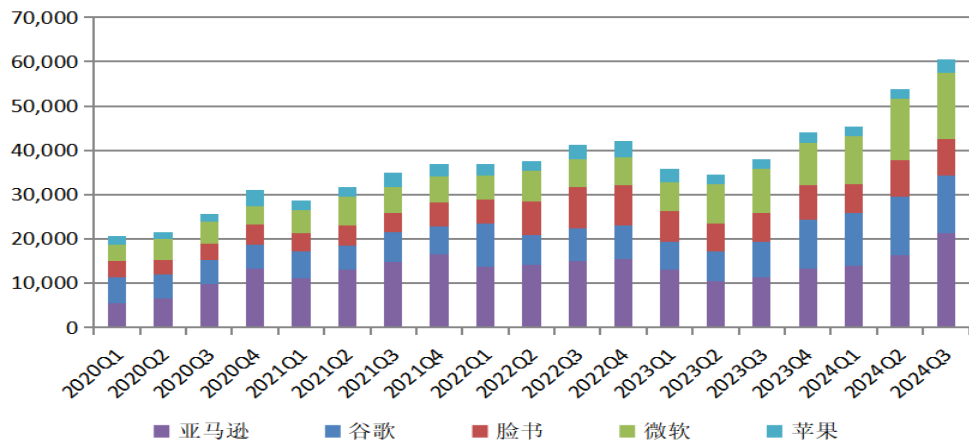
(2) 英伟达主推的新一代计算单元 GB200 NVL72，单机柜功率预计将达到 125KW，而多数数据中心部署的标准机架仅为约 12kW 至 20kW，这意味着在散热、供电、网络设计、并行、可靠性等方面都将迎来大量的挑战，也将有大量数据中心开启液冷方案改造。

从最新的内存发布进度来看，三星、SK 海力士与美光已分别于 2024 年上半年和第三季提交首批 HBM3e 12hi 样品，目前处于持续验证阶段。其中 SK 海力士与美光进度较快，有望于今年底完成验证。

3.2. 海内外互联网厂商的资本投入创新高

国外来看，亚马逊、谷歌、苹果、微软、脸书为代表的 5 大科技厂商的总资本开支继续呈现提升趋势，已经在 2024Q3 达到了 604.28 亿美元，同比增长了 59%。其中亚马逊领涨，达到 88%，总金额也位列第一；谷歌以 62% 的增速，紧追微软；微软以 50% 的增速，总资本开支金额位列第二；Meta 和苹果资本开支分别增长 26% 和 34%，总资本开支金额也远不及其他厂商。

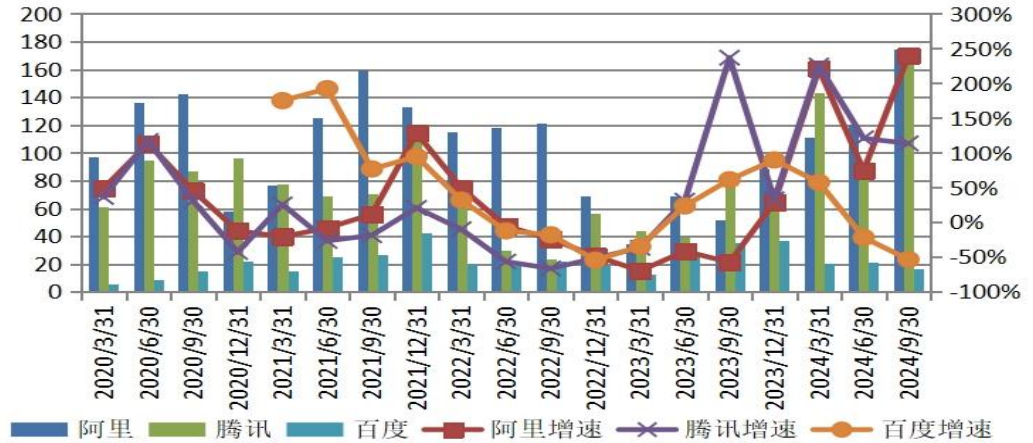
图 25：20Q1-24Q3 5 大科技厂商资本投入（百万美元）



资料来源：上市公司财报，Wind，中原证券研究所

国内互联网厂商的资本开支在 Q2 环比走弱以后，Q3 呈现明显加快趋势，创出历史新高。3 家资本开支总计 362.22 亿元，同比增长 117.2%。其中阿里资本支出 174.91 亿元，同比增速达到了 239.6%，达到了历史新高；腾讯资本开支 170.94 亿元，同比增速 113.5%，也再创新高；但百度 Q3 资本开支仅为 16.37 亿元，同比下滑了 53.6%，连续两个季度下滑。

图 26：2020Q1-2024Q3BAT 的资本开支（亿元）



资料来源：上市公司财报，中原证券研究所

互联网企业的资本开支加大，带来了上游对服务器需求的结构性变化，这其中：

(1) 对标海外市场的厂商中，工业富联前三季度云计算业务同比增长 71%，其中 AI 服务器收入同比增长 228%，联想集团基础设施解决方案业务上半年同比增长 44%。

(2) 国内互联网企业的重要供应商中，浪潮信息前三季度同比 72%，但是毛利率从去年的 9.67% 下滑到 6.70%（互联网厂商采购量大，毛利率更低）；受到互联网投资力度加大，白牌服务器订单增加等因素影响，新华三政企业务前三季度同比增长 11.91%。

表 1：企业服务器相关业务数据对比（亿元）

公司	相关业务	2024H1						2024.1-9			
		收入	增速	净利润	增速	毛利率	毛利	收入	增速	收入两年增速	毛利两年增速
工业富联	云计算		60%	-	-				71%		
	-AI 服务器		230%						228%		
浪潮信息	-	421	69%	6.0	91%	7.7%	33	831	72%	58%	-12%
联想集团	基础设施解决方案	406	44%								
紫光股份	-	380	5%	10	-2%	19%	72	588	7%	9%	-9%
	ICT 基础设施及服务	262	5%			24%	64				
中科曙光	新华三	264	6%	18	0%			403	7%	10%	
	-	57	6%	5.6	3%	26%	15	80	4%	9%	21%
软通动力	IT 设备业务	51	4%	-	-	14%	6				
	计算产品与数字基础设施服务	39				6%	2.5				
神州数码	同方计算机	15		-0.67							
	自主品牌业务(旧)	23.7	57%	-	-						
神州数码	自有品牌产品(新)	26.6	46%	0.44		10.5%					
	-信创业务	21.9	73%			12.1%					
拓维信息	-神州泰坤人工智能服务器	5.6	273%								
	计算机、通信和其他电子设备制造业	8.7	89%			10%	0.9				
拓维信息	湘江鲲鹏	7.0	243%	0.007	100%						

资料来源：上市公司公告，Wind，中原证券研究所

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/928107103116007006>