

证券研究报告

2024年09月01日

行业报告：行业深度研究

计算机

# 国产算力B端篇：云厂商25年有望开启规模采购

作者：

分析师 缪欣君 SAC执业证书编号：S1110517080003

联系人 刘鉴



天风证券  
TF SECURITIES

行业评级：强于大市（维持评级）  
上次评级：强于大市

请务必阅读正文之后的信息披露和免责声明

# 摘要

## 1、AI基础设施推动24Q2 BAT资本开支单季度同比68%，24年增速预计超57%

根据BAT互联网大厂2024年Q2季报及电话会议，百度Q2资本开支21.18亿元，同比-22%；阿里巴巴Q2资本支出119.39亿元，同比73%，电话会表示AI产品在训练、推理，以及像通义千问 API 调用需求都在快速增长；腾讯Q2资本87.29亿元人民币，同比118%，或投入AI领域。结合VisibleAlpha一致预期，我们预计2024年BAT总资本支出约1090.95亿元，同比增速57.1%。

## 2、训练侧快速迭代，推理侧积极探索AIGC APP或嵌入式AI等新应用场景，未来算力需求有望进一步提升

训练侧，国产大模型在过去一年快速迭代，目前阶跃星辰、MiniMax等厂商已发布万亿参数大模型，未来或向十万亿参数继续迭代，带动训练算力需求增长。推理侧，根据QuestMobile公众号数据，截止到6月份，月活千万级以上用户的APP数量为2个，豆包、文心一言分别为2752万、1134万。我们认为，以豆包、文心一言为首的AIGC APP 6月月活增速较快，虽然用户活跃率仍待提高，但我们认为互联网厂商仍然通过推广不断拉新。同时，互联网厂商不断尝试如智能体、嵌入式AI等新应用场景。我们认为，训练和推理侧的共同发力有望推动算力需求进一步提升。

## 3、政策引领国产算力景气度提升，25年英伟达H20增量/存量数据中心或适用阶梯电价

2024年7月23日，国家发改委等部门发布关于印发《数据中心绿色低碳发展专项行动计划》的通知，新建及改扩建数据中心应采用能效达到所提标准文件规定的节能水平及以上服务器产品，即14nm以下制程芯片须 $\geq 0.5\text{Tflops/W}$ ，25年英伟达H20增量/存量数据中心或适用阶梯电价，国产芯片竞争力有望进一步提升。

## 4、BAT资本支出未来4年CAGR预计26.31%，有望加速向国产算力转换

目前微软、谷歌、Meta、亚马逊等海外互联网公司24年资本支出有望超2000亿美元，增速44.10%，占收入比重近17%。我们认为，随着国内大模型的迭代，未来BAT资本支出占比有望向海外厂商靠拢。预计27年BAT资本开支将达到1767.44亿元，23-27年资本开支复合增速26.31%。

### 建议关注：

- 1) 四小龙：寒武纪、海光信息、神州数码、中科曙光
- 2) 华为：软通动力、烽火通信、广电运通、拓维信息
- 3) 英伟达：浪潮信息、智微智能（与通信组联合覆盖）

**风险提示：** AI算力景气度下降的风险、AI芯片竞争加剧的风险、政策落地不及预期风险

# 2023年GPU占据85%市场份额，互联网仍为主要需求方，占比近60%

- ◆ **AI芯片：**根据IDC数据，2023年中国加速芯片的市场规模达到近140万张。从技术角度来看，GPU卡占据85%的市场份额
- ◆ **行业：**根据IDC数据，2023年互联网是最大的采购行业，占整体加速服务器市场近60%的份额，此外金融、电信、交通和医疗健康等大多数行业均有超过一倍以上的增涨。

图1：2023年作为AI元年，GPU卡占据85%市场份额，约119万张

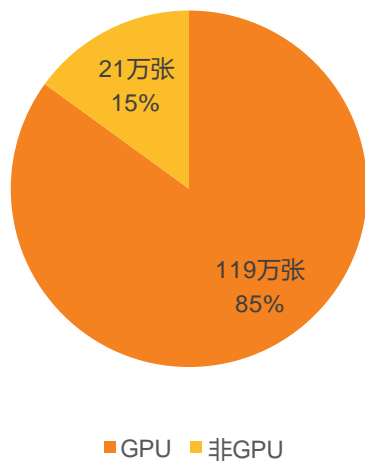
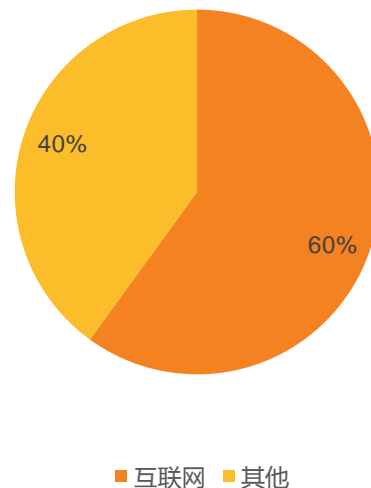


图2：互联网是智算需求最大的行业，占比近60%



# 大模型推动下互联网资本开支24Q2同比增加68%，多数投入AI及算力基础设施

根据BAT互联网大厂季报及电话会议，2024年Q2资本开支大幅增加，同比68%，多数投向AI及算力基础设施：

- ◆ **百度**：24Q2资本开支21.18亿元人民币，同比降低22%，百度在业绩电话会表示，正在加速搜索业务的AI转型，通过引入AI生成的搜索结果和交互功能，以提供更丰富和个性化的用户体验。
- ◆ **阿里巴巴**：24Q2资本性支出119.39亿元人民币，同比73%，业绩电话会白表示，客户的需求非常旺盛，包括客户的pipeline现在还远远没有到满足的阶段。从客户对于AI产品在训练、推理，以及像通义千问API调用，这方面的需求都在快速增长。
- ◆ **腾讯**：24Q2资本开支87.29亿元人民币，同比118%，或投入AI领域。

表1：BAT24Q2资本开支主要投向AI及算力基础设施（亿元，人民币）

	24Q2	YOY	资本开支主要投向
百度	21	-22%	正在加速搜索业务的AI转型，引入AI生成的搜索结果和交互功能
阿里巴巴	119	73%	AI产品在训练、推理，以及像通义千问API调用需求都在快速增长
腾讯	87	118%	或投入AI领域
合计	228	68%	

# 训练侧：国产大模型版本快速迭代，带动训练算力需求增长

- ◆ 百度在一年多的时间里不断扩展其模型阵容，连续更新其大型模型。继24年4月推出文心4.0工具版后，5月又推出了性能更优的文心4.0 Turbo。目前，百度仅用四年时间完成文心大模型六次迭代。
- ◆ 自去年4月通义千问一年内多次迭代，24年5月9日阿里云发布通义千问2.5，开源1100亿参数模型。
- ◆ 腾讯混元大模型经过近一年的迭代升级，在国内率先采用MoE架构，参数量已达到万亿，tokens数量超过7万亿。
- ◆ 24年5月15日，经过一年时间的迭代和市场验证，字节自研大模型豆包大模型正成为国内使用量最大、应用场景最丰富的大模型之一，目前日均处理 1200 亿 Tokens 文本，生成 3000 万张图片。
- ◆ 目前，阶跃星辰、MiniMax等模型厂商已发布万亿参数模型，未来或向十万亿参数继续迭代，训练算力需求有望进一步提升。

图3：豆包大模型正成为国内应用场景最丰富的大模型之一



图4：通义千问一年时间快速迭代，24年5月发布2.5版本



# 推理侧：豆包、文心一言月活突破千万，各大厂商正在积极投流

- ◆ 根据QuestMobile数据，截止到6月份，月活千万级以上用户的APP数量为2个，豆包、文心一言分别为2752万、1134万。五百万至千万的APP仅1个，百万至五百万的APP为7个。至于大部分玩家（63个），月活用户在50万以下，行业占比超8成。当前主流AIGC APP的月人均使用天数基本低于5天，导致活跃率普遍低于15%。
- ◆ 从流量构成来看，纯拉新用户占主要份额（与成长初期有关），尤其是豆包、通义千问。23年8月豆包发布，至24年6月仅10个月时间用户量已达2751.5万。我们认为，从大模型活跃度和用户量快速增长来看，互联网厂商正在努力尝试AI新概念，未来有望继续投流。

图5：豆包上线仅10个月月活突破2700万



注：1. QuestMobile TRUTH数据库收录标准：MAU>1万；2. 月复合增长率=(期末月用户规模/起始月用户规模)^(1/时间段)-1。针对特定APP，如2023年6月未被QM收录，则按照实际收录月份作为起始值，时间段进行相应调整



注：1. QuestMobile TRUTH数据库收录标准：MAU>1万；2. 月复合增长率=(期末月用户规模/起始月用户规模)^(1/时间段)-1。针对特定APP，如2023年6月未被QM收录，则按照实际收录月份作为起始值，时间段进行相应调整

图6：从24年6月数据看，主流AIGC APP用户活跃率较低

AIGC APP	活跃率 (DAU/MAU)	月人均使用天数 (天)	月卸载率
豆包	15.1%	4.5	49.5%
文心一言	16.8%	5.0	32.0%
天工	15.2%	4.6	36.8%
Kimi智能助手	14.1%	4.2	33.6%
讯飞星火	13.3%	4.0	38.0%
通义	14.3%	4.3	62.0%
海螺AI	12.4%	3.7	47.4%
星绘	7.5%	2.3	92.0%
腾讯元宝	10.7%	3.2	53.3%

注：1. 依据APP月活跃用户规模降序排序；2. 活跃率：在统计周期(周/月)内，该App的日均活跃用户数与其月活跃用户数的比值；3. 人均使用天数：在统计周期内，平均每个人使用APP的天数

# 推理侧：厂商积极探索嵌入式AI等新场景，有望拉动推理算力需求进一步提升

- ◆ **智能体概念出现，厂商积极探索：**今年1-6月，头部APP均有多次版本更新迭代，除模型能力上的提升外，以用户视角嵌入更多多元化的场景功能，从而提升使用体验，带动用户粘性。
- ◆ **嵌入式AI成为新战场：**QuestMobile数据显示，支付宝AI金融助手6月月活用户超5900万，且APP首页新增生活助理入口。电商领域，智能导购如"淘宝问问"和"京东京言"助力用户购物决策。抖音电商测试"AI购物小助手"，通过对话推荐商品。钉钉升级推出AI搜索、助理、会议等多Agent功能，日调用量破千万。
- ◆ 我们认为，以豆包、文心一言为首的AIGC APP月活增速较快，虽然或主流AIGC APP用户活跃率较低但已表明互联网及大模型厂商通过推广不断拉新。此外，大模型厂商不断尝试如智能体、嵌入式AI等新应用场景，有望拉动推理算力需求进一步提升。

图7：智能体概念被推向台前

AIGC APP	初始智能体个数	细分场景举例
豆包	8个	豆包   英语外教Owen   英语学习助手   学习小帮手   全能写作助手   AI图片生成   超爱聊天的小宁   编程助理
天工	10+个	天工   AI文档-音频分析师   AI写作   AI文档分析   AI图片生成   旅游规划师   二次元漫画   豆瓣推荐   星座运势...
讯飞星火	3个	小说创作   大学专业怎么选   健康问题百科大全
海螺AI	4个	帮我识图   帮我写作   帮我做攻略   和我语音通话



图8：内嵌式AI潜力初现



# 缺货加禁令影响下互联网“芯荒”，国产芯片或迎新曙光

- ◆ 2023年以来，以Transformer结构为主的生成式带动了加速计算的需求突破式增长，随着Scalinglaw的进一步拓展，英伟达GPU极度短缺。
- ◆ 2023年10月17日，美国发布了针对芯片的出口禁令新规，对于中国半导体的制裁进一步升级，禁令下国内互联网厂商面临“芯荒”风险。

## 国产芯片或迎新曙光：

- ◆ 根据维科网，百度为200台服务器订购了1600个华为910B昇腾人工智能芯片，总值约4.5亿元；
- ◆ 根据快科技报道，未来字节跳动、腾讯、阿里巴巴和百度等大科技公司有望减少购买外国制造的AI芯片，转而购买更多国产芯片，国产芯片或迎新曙光。

图9：ScalingLaw下随大模型参数量增长算力缺口逐步扩大

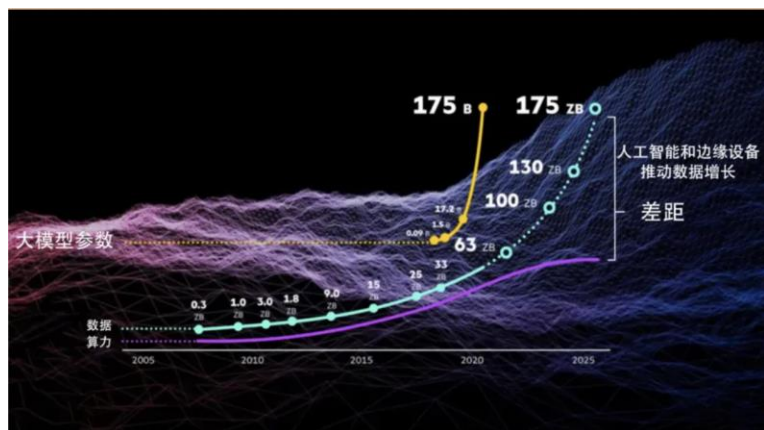


图10：缺货+禁令双重因素致互联网“芯荒”



FOR IMMEDIATE RELEASE  
October 17, 2023  
<https://bis.doc.gov>

BUREAU OF INDUSTRY AND SECURITY  
Office of Congressional and Public Affairs  
Media Contact: [OCPA@bis.doc.gov](mailto:OCPA@bis.doc.gov)

Commerce Strengthens Restrictions on Advanced Computing Semiconductors, Semiconductor Manufacturing Equipment, and Supercomputing Items to Countries of Concern

Updates to Modify and Reinforce Restrictions Initially Released on October 7, 2022, to Address National Security Concerns Posed by PRC Military Modernization



# 政策引领国产算力景气度提升，25年英伟达H20增量/存量数据中心或适用阶梯电价

7月23日，国家发改委等部门发布关于印发《数据中心绿色低碳发展专项行动计划》的通知，提到：

- ◆ **目标：**到2025年底，全国数据中心布局更加合理，整体上架率不低于60%，平均电能利用效率降至1.5以下。市域内已有建成投用1年以上数据中心，且整体上架率低于50%的，原则上不再规划建设新的数据中心集群及大型和超大型数据中心项目。
- ◆ **要求：**新建及改扩建数据中心应采用能效达到《塔式和机架式服务器能效限定值及能效等级》（GB43630-2023）和《服务器和数据存储设备能效“领跑者”评价要求》（T/CECA-G0284-2024）规定的节能水平及以上服务器产品。
- ◆ **电价优惠政策：**全面清理各地区自行出台的高耗能行业电价优惠政策，推动建立统一的工业重点领域阶梯电价政策。
- ◆ **利好国产算力：**从政策看，25年英伟达H20增量/存量数据中心或适用阶梯电价，国产芯片竞争力有望进一步提升

表3：H20不满足节能水平要求

品牌	型号	峰值性能 (FP16, TFlops)	功耗 (W)	能效系数 (TFlops/W)
英伟达	H20	148	400	0.37
昇腾	910	320	310	1.03
寒武纪	MLU370-X4	96	150	0.64
摩尔线程	S4000	100	450	0.22
燧原	云燧i20	128 (邃思2.5性能)	150	0.85
沐曦	曦思N100	80	70	1.14
天数	天垓100	128	250	0.51
昆仑芯	R200	128	150	0.85

表2：规定配置小于14nm设备制造的通用图形处理器节能水平应  $\geq 0.5$ TFlops/W（FP64大于2TFlops的AI芯片不受文件限制）

单位为每秒浮点运算次数每瓦 (TFLOPS/W)

通用图形处理器类型	通用图形处理器能效	
	先进水平	节能水平
配置大于或等于 14 nm 设备制造的通用图形处理器，或配置非 EUV 设备制造的通用图形处理器	$\geq 0.50$	$\geq 0.25$
配置小于 14 nm 设备制造的通用图形处理器	$\geq 1.00$	$\geq 0.50$

# 海外互联网厂商：24年资本支出增速超44%，主要集中在算力基础设施

- ◆ 资本支出大幅增加，集中在算力基础设施：四家科技公司在今年第一季度业绩报告会上表示，仍将大幅增加资本支出，集中在数据中心、芯片、以及其他用于构建或部署生成式AI模型的设备。

图11：微软24年资本支出占收入比重预计超18%（亿美元）

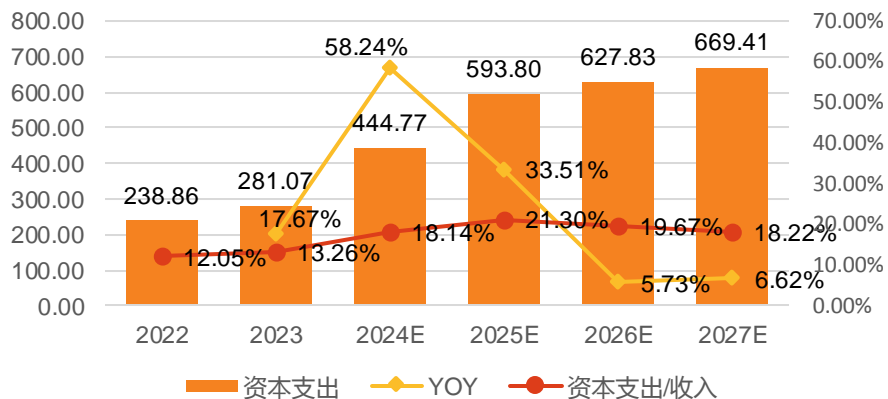


图12：谷歌24年资本支出占收入比重预计同比4.15pct（亿美元）

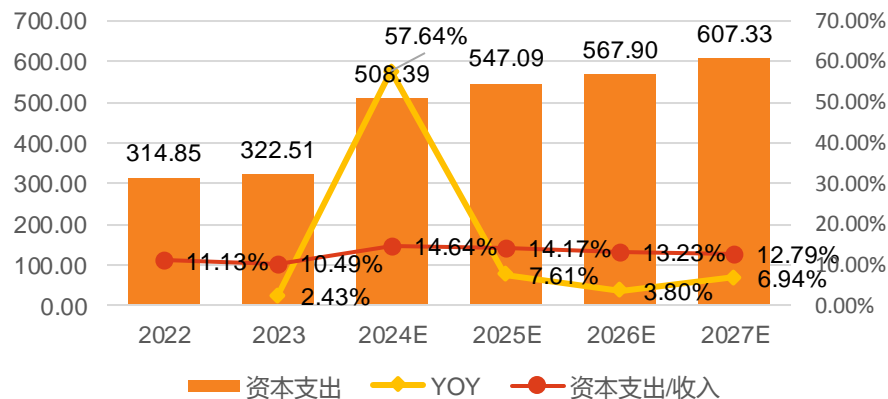


图13：Meta 24年资本支出预计同比超40%（亿美元）

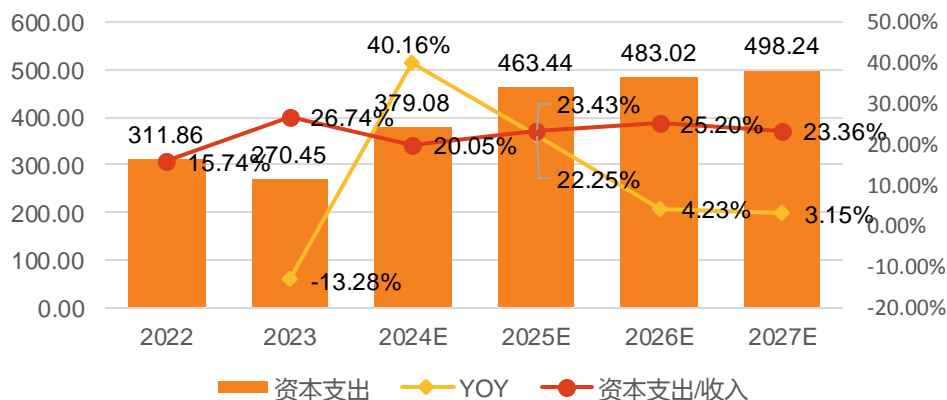
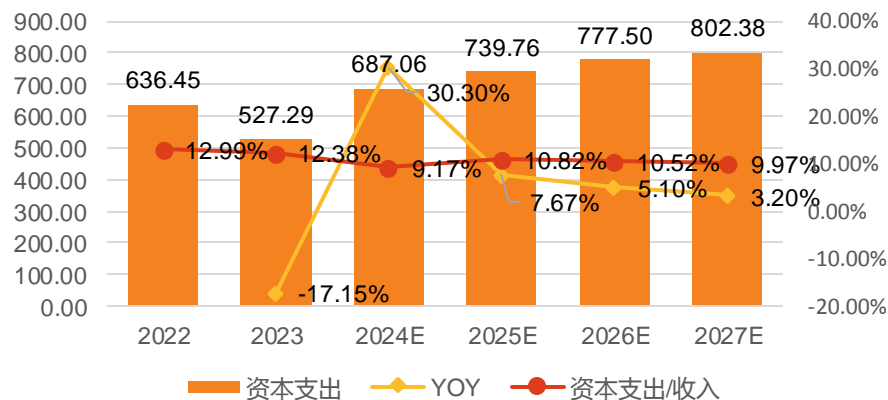


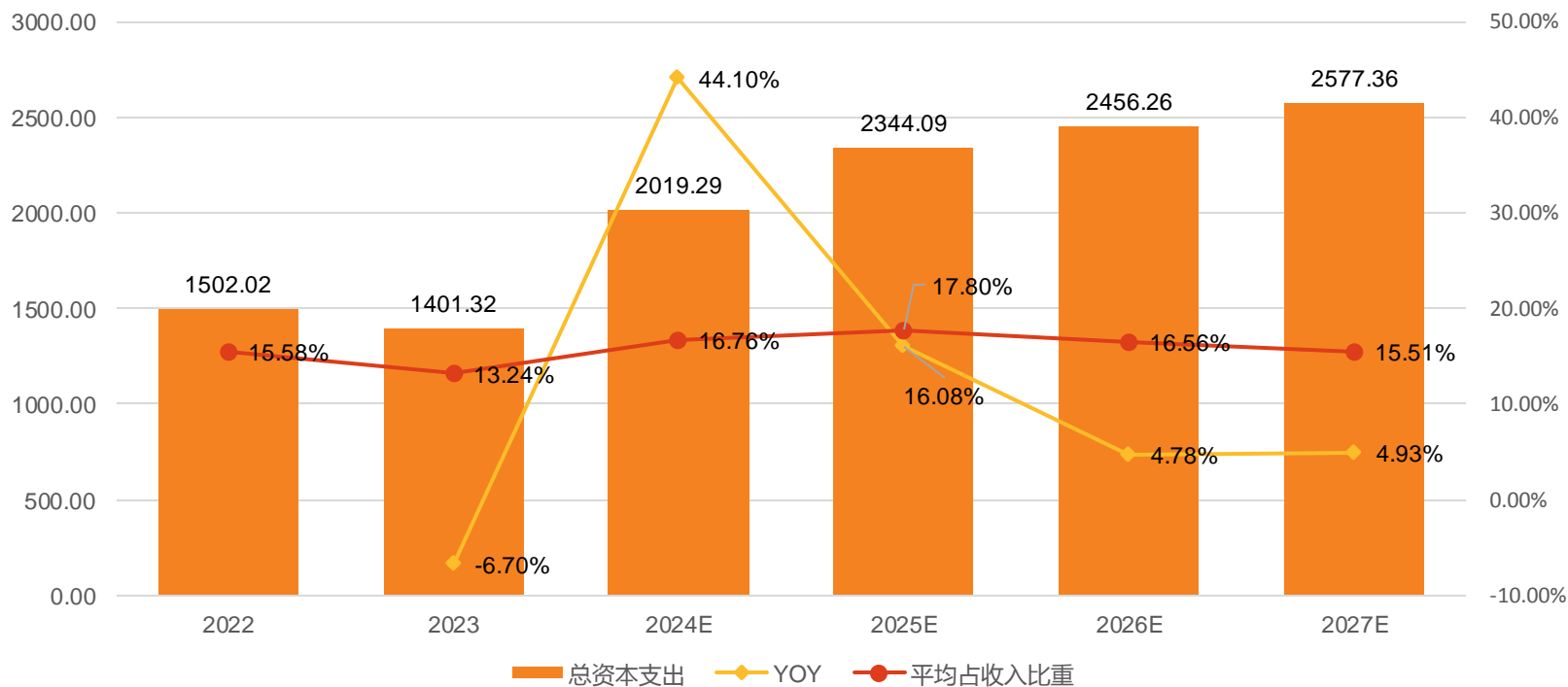
图14：亚马逊24年资本支出占收入约30.3%，25年以后增速有所放缓（亿美元）



# 海外互联网厂商：24年预计资本支出平均占收入比重近17%，远超国内

- ◆ 海外互联网资本支出占收入比重持续提升：根据VisibleAlpha一致预期，2024年微软、谷歌、Meta、亚马逊资本支出预计2019.29亿美元，同比增加44.10%；资本开支占收入比重预计达16.76%，对比去年同比增加3.51pct。

图15：2024年海外互联网资本支出预计超2000亿美元，增速超44%，占收入比重近17%（亿美元）



# 国内互联网厂商：24年仅BAT三家资本支出或超1090亿元，同比57.1%

- ◆ **BAT资本支出近两年预计保持高增长：**百度、阿里巴巴、腾讯三家互联网公司资本支出逐步提升，24Q2三家总资本支出对比23年同期同比68%，根据各公司财报与电话会议，资本开支主要投向AI及算力基础设施。我们认为，随着国内大模型的迭代，未来国内BAT资本支出占比有望向海外厂商靠拢。

图16：24-25年阿里巴巴资本支出预计高增（亿元）

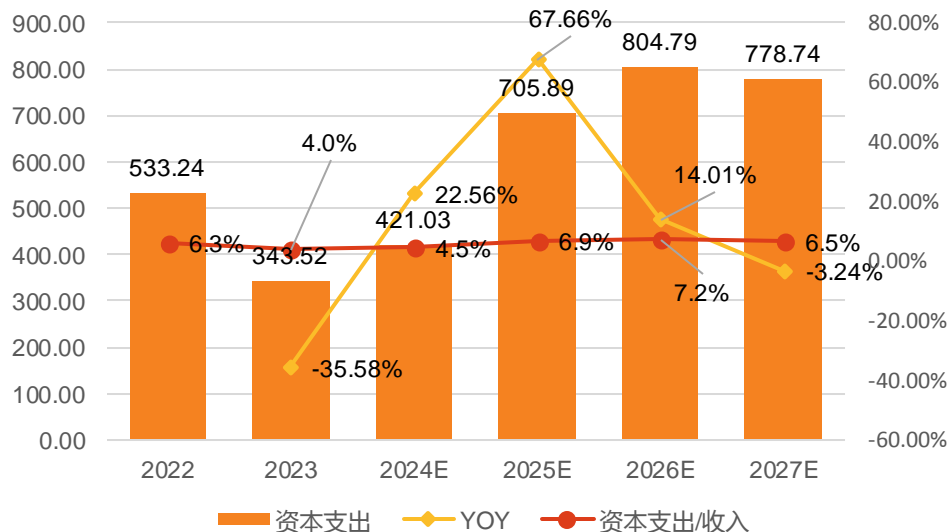


图17：腾讯24年资本支出预计翻倍（亿元）

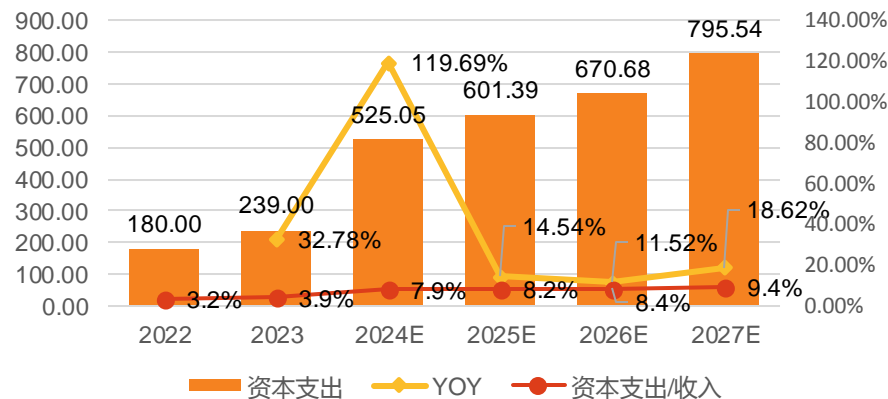
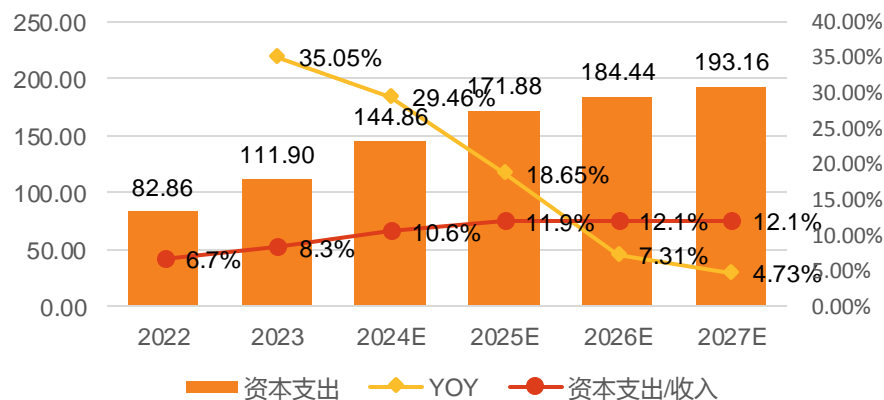


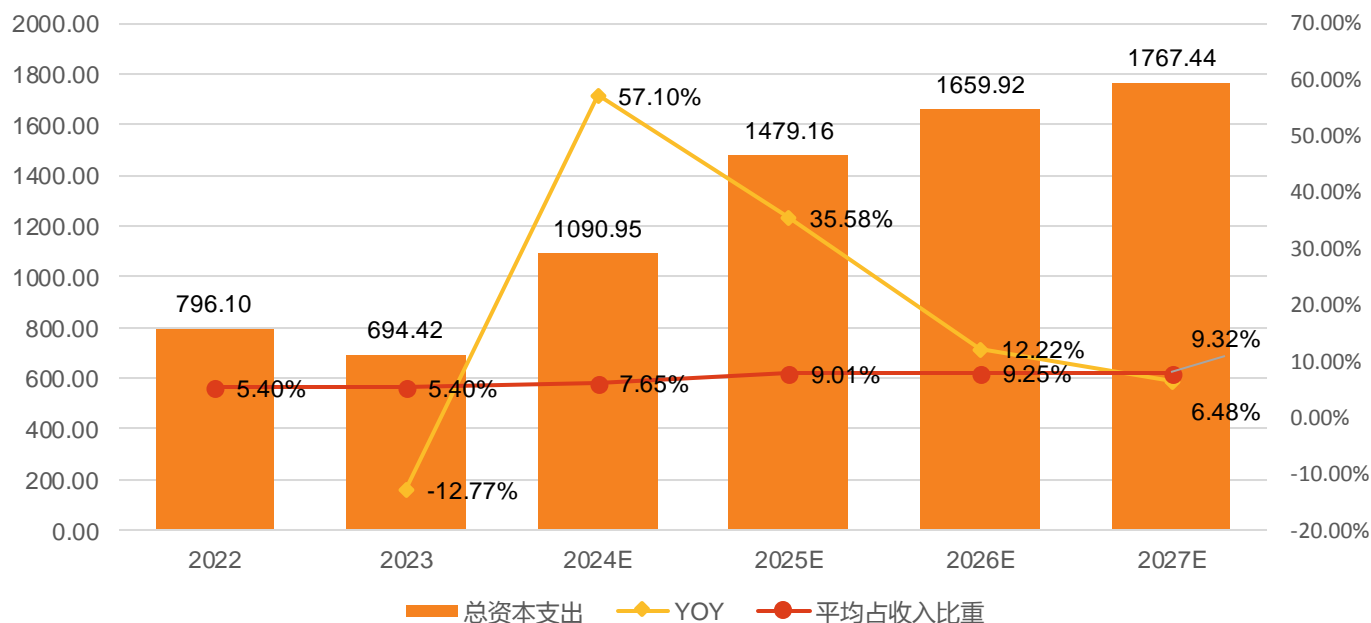
图18：百度资本支出增速有所放缓，或与去年投入较多有关（亿元）



## 国内互联网厂商：未来4年BAT资本开支CAGR 26.31%，有望加速向国产算力转换

- ◆ **算力需求进一步提升：**2023年以来，以Transformer结构为主的生成式带动了加速计算的需求突破式增长，随着模型参数量向万亿发展，我们认为，未来训练+推理算力需求有望持续提升，占资本支出比重有望进一步增加。
- ◆ **资本支出近两年保持高增长：**百度、阿里巴巴、腾讯三家互联网公司资本支出逐步提升，24Q2三家总资本支出对比23年同期同比68%，根据各公司财报与电话会议，资本开支主要投向AI及算力基础设施。目前微软、谷歌、Meta、亚马逊等海外互联网公司资本支出占收入比重持续增加，23年已达13.24%。我们认为，随着国内大模型的迭代，未来国内BAT资本支出占比有望向海外厂商靠拢。
- ◆ **国产算力比例有望持续提升：**根据《数据中心绿色低碳发展专项行动计划》，25年英伟达H20增量/存量数据中心或适用阶梯电价，国产AI芯片竞争力有望进一步提升。

图19：考虑算力投入，预计BAT资本支出占收入比重与海外互联网公司逐步缩小差距（亿元）



## 建议关注与风险提示

### 建议关注：

- 1) **四小龙**：寒武纪、海光信息、神州数码、中科曙光
- 2) **华为**：软通动力、烽火通信、广电运通、拓维信息
- 3) **英伟达**：浪潮信息、智微智能（与通信组联合覆盖）

### 风险提示：

**AI算力景气度下降的风险**：算力支出与下游应用息息相关，若AI应用需要更长期才能突破，则算力支出的高景气可能不可持续

**AI芯片竞争加剧的风险**：AI芯片领域有较多参与者，未来市场竞争可能加剧

**政策落地不及预期风险**：地方政府智算中心主要依靠各地政策推动，若政策落地不及预期，则可能影响智算中心算力建设相关公司

## 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，本报告所表述的所有观点均准确地反映了我们对标的证券和发行人的个人看法。我们所得报酬的任何部分不曾与，不与，也将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。

## 一般声明

除非另有规定，本报告中的所有材料版权均属天风证券股份有限公司（已获中国证监会许可的证券投资咨询业务资格）及其附属机构（以下统称“天风证券”）。未经天风证券事先书面授权，不得以任何方式修改、发送或者复制本报告及其所包含的材料、内容。所有本报告中使用的商标、服务标识及标记均为天风证券的商标、服务标识及标记。

本报告是机密的，仅供我们的客户使用，天风证券不因收件人收到本报告而视其为天风证券的客户。本报告中的信息均来源于我们认为可靠的已公开资料，但天风证券对这些信息的准确性及完整性不作任何保证。本报告中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，天风证券及/或其关联人员均不承担任何法律责任。

本报告所载的意见、评估及预测仅为本报告出具日的观点和判断。该等意见、评估及预测无需通知即可随时更改。过往的表现亦不应作为日后表现的预示和担保。在不同时期，天风证券可能会发出与本报告所载意见、评估及预测不一致的研究报告。

天风证券的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。天风证券没有将此意见及建议向报告所有接收者进行更新的义务。天风证券的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

## 特别声明

在法律许可的情况下，天风证券可能会持有本报告中提及公司所发行的证券并进行交易，也可能为这些公司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。因此，投资者应当考虑到天风证券及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突，投资者请勿将本报告视为投资或其他决定的唯一参考依据。

## 投资评级声明

类别	说明	评级	体系
股票投资评级	自报告日后的6个月内，相对同期沪深300指数的涨跌幅	买入	预期股价相对收益20%以上
		增持	预期股价相对收益10%-20%
		持有	预期股价相对收益-10%-10%
		卖出	预期股价相对收益-10%以下
行业投资评级	自报告日后的6个月内，相对同期沪深300指数的涨跌幅	强于大市	预期行业指数涨幅5%以上
		中性	预期行业指数涨幅-5%-5%
		弱于大市	预期行业指数涨幅-5%以下

THANKS